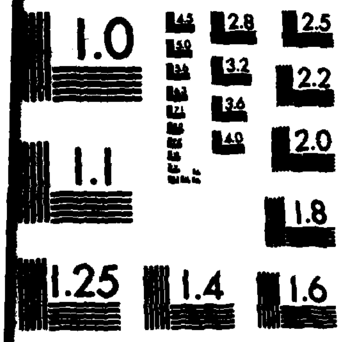
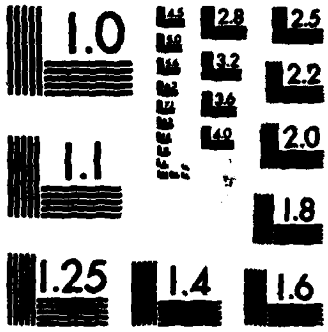
AD-A120 687 A MIXING DISTRIBUTION APPROACH TO ESTIMATING PARTICLE SIZE DISTRIBUTIONS(U) STANFORD UNIV CA DEPT OF STATISTICS A V KUK 19 OCT 82 TR-328 N00014-76-C-0475 1/2

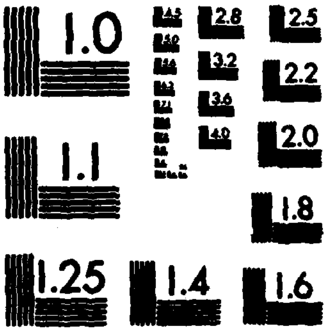
UNCLASSIFIED F/G 12/1 NL



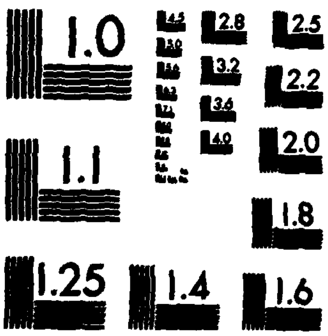
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



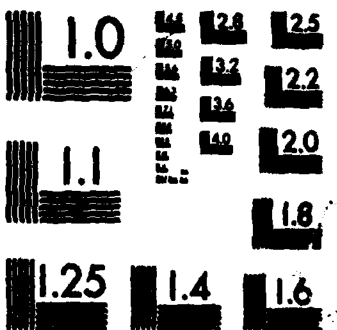
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A MIXING DISTRIBUTION APPROACH TO ESTIMATING PARTICLE SIZE DISTRIBUTIONS

By

Anthony Yung C. Kuk

TECHNICAL REPORT NO. 328

October 19, 1982

Prepared Under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ACKNOWLEDGEMENTS

It is a pleasure to thank my advisor, Professor Herbert Solomon, for introducing me to the field of stereology and for his technical assistance and personal support. To Professor Paul Switzer who has been most encouraging, I owe sincere gratitude. I would like to thank Professor Bradley Efron for reading this thesis.

I benefited from helpful discussion with Professor Adrian Baddeley while he was visiting Stanford. I am also grateful to Professor R.S. Anderssen for his listings of programs on spectral differentiation. The assistance of Stephen Tsun in the areas of computing and writing is appreciated. I thank and admire Carolyn Knutsen for her patience and good cheer in typing the numerous drafts of the thesis.

This thesis would not have been possible without the constant support, encouragement and showing of understanding from concerned friends, especially Nelson Chan. Helen Chang, Peter Fu, Max Hui and Florence Tam. I am greatly indebted to my sister and brother-in-law for putting me through college and to my dear parents for their love. Finally, my experience at Stanford is enriched beyond measure because of Christ who has come into my life. To Him, I owe the greatest gratitude.

TABLE OF CONTENTS

		Page
INTRODUCTION		1
CHAPTER I	GENERAL THEORY	4
1.1	Description of the Stereological Problem	4
1.2	The Classical Procedure	5
	1.2a Random Line	5
	1.2b Random Plane	5
	1.2c Thin Slice of Thickness 2τ	6
1.3	The Basic Problem with the Classical Procedure	7
1.4	Formulation of the Problem of Estimating Mixing Distributions	8
1.5	The Connection Between the Stereological Problem and Estimation of Mixing Distributions	9
	1.5a Random Line	10
	1.5b Random Plane	11
	1.5c Thin Slice of Thickness 2τ	12
1.6	Summary of Notations	14
1.7	The New Approach	15
1.8	Nonparametric Estimation of Mixing Distributions	16
	1.8a Nonparametric Maximum Likelihood Estimate (MLE)	16
	1.8b Minimum Distance Method	19
1.9	The Advantage of the New Approach Over the Classical Approach	20
1.10	Some Peculiar Features	21
CHAPTER II	THE RANDOM LINE CASE	23
2.1	The Basic Formulas	23
2.2	Survey of the Literature	24

	Page
2.3 Nonparametric MLE	25
2.3a Derivation and Computation	25
2.3b Argument for Using MLE	27
2.3c Consistency of MLE	27
2.4 Simulation Results	27
2.5 Truncation	34
CHAPTER III THE RANDOM PLANE CASE	34
3.1 The Basic Formulas	34
3.2 Survey of the Literature	36
3.3 Characterization of $h(\cdot \mid \uparrow)$	37
3.4 Nonparametric MLE	39
3.4a Derivation	39
3.4b Inconsistency of MLE	43
3.5 Minimum Distance Methods	44
3.5a Description	45
3.5b Consistency of $\hat{G}_n(\theta \mid \uparrow)$	45
3.5c A Difficulty in Proving Consistency of $\hat{G}_n$	47
3.5d Modification Leading to Consistent Estimate of G_0	48
3.5e Comparison With Existing Methods	50
3.6 Simulation Results	50
3.7 Bootstrap	58
3.7a Why is it Necessary	58
3.7b Introduction	59
3.7c The Wrong Way to Bootstrap	61
3.7d The Right Way to Bootstrap	61

Page

3.7e Simulation Results 62

3.7f Conclusions and Applications 72

3.8 Truncation 73

CHAPTER IV THIN SLICE OF THICKNESS 2τ 74

4.1 The Basic Formulas 74

4.2 Decomposition and Inversion 74

4.2a G Continuous 74

4.2b A Basic Fact 75

4.2c G mixed 76

4.2d G Discrete 77

4.3 Decomposition of Data, H_n and F_{Q*} 80

4.3a Notations Concering the Data 80

4.3b Decomposition of Data 80

4.3c Decomposition of H_n 81

4.3d Decomposition of F_{Q*} 83

4.4 Methods Based on Inversion Formulas 85

4.4a G Continuous 85

4.4b G Mixed 85

4.4c G Discrete 85

4.5 A Closer Look at the Discrete Case 86

4.5a Treating Part of Thin Slice Data as Planar Data 86

4.5b A Basic Equation 86

4.5c Relationship Between Φ_n and Θ 87

4.5d G Discrete, m Finite - Further Reduction 89

4.6 Miscellaneous Lemmas 90

	Page
4.7 Minimum Distance Method	91
4.7a Description	91
4.7b Consistency of $\hat{Q}_n$	92
4.7c Consistency of $\hat{G}_n$	95
4.7d Discrete Case - Further Simplification	96
4.8 Simulation Results	96
APPENDIX	96
A.1	96
A.2 Proof of Lemmas 3.5.1 and 3.5.2	109
A.3 Proofs of the Lemmas of Section 4.6	115
REFERENCES	122

INTRODUCTION

Unfolding (estimating) a particle size distribution is an old and well known problem in stereology which has not yet been solved satisfactorily. A description of the problem is as follows: spherical particles are dispersed randomly in a three-dimensional body. The classical assumption is that the centers of the spheres are distributed according to a dilute Poisson process. The radii of such spheres have a distribution G independent of everything else. A random probe (line, plane, or thin slice) is cut through the volume. The observations y are then the following:

i) For the random line, its intersection with a sphere is a line segment, y is half the length of that segment.

ii) For the random plane, y is the radius of the circle of intersection.

iii) For the thin slice, y is the profile radius, that is, the maximum of the radii of the circles of intersection. The problem is to estimate the distribution G from the observations $y_1,\ldots,y_n$.

Past procedures make use of the inversion formula that expresses the particle size distribution G as a function of the distribution H of the observed data. An estimate $\tilde{G}$ is obtained by replacing H in the formula by an estimate $\tilde{H}$ where $\tilde{H}$ is either the sample c.d.f. or a smoothed version of it. These procedures do not take the structure of the problem into account. Consequently, they have the following shortcomings:

i) $\tilde{H}$ does not belong to the admissible range of H with the result that $\tilde{G}$ is not a distribution function.

ii) The same estimate of H is used regardless of the kind of data (linear, planar or thin slice).

iii) They cannot incorporate additional information concerning G.

In this dissertation, we propose a new procedure from the viewpoint of nonparametric estimation of mixing distributions (see sections 1.4, 1.5, 1.7). In section 1.9, we show how the new approach can deal with all three shortcomings of the classical procedures. We consider linear, planar, and thin slice data (chapters II, III and IV respectively). In all three cases, our approach performs better than the classical procedures (sections 2.4, 3.6, 4.8). In addition, we prove consistency results (sections 2.3, 3.5, 4.7).

In the random line case, the nonparametric maximum likelihood estimate (MLE) of G is derived (section 2.3). However, because of a "peaking" effect of the MLE near the mode, we obtain extremely poor simulation results. By modifying the procedure slightly, we are able to obtain much better results (section 2.4).

The importance of having $\tilde{H}$ lie within the admissible range is highlighted in the random plane case where we want to bootstrap the distribution of a stereological estimate. In section 3.7, we show that it is crucial to resample from an estimate $\tilde{H}$ that belongs to the admissible range, since the sample c.d.f. does not, the usual way of resampling from the sample c.d.f. will not work. We also indicate how bootstrap methods can contribute to stereology.

In the thin slice case, the formulas derived in the literature are valid only when G is continuous. In section 4.2, we derive the correct formulas when G is mixed or discrete. These formulas involve a decomposition of H into its continuous and discrete component. This makes the estimation problem more complicated but also more interesting especially in the discrete case. We propose a few procedures which involve a decomposition of the data corresponding to that of H (sections 4.3, 4.5, 4.7). The discrete case is more interesting because of the following:

i) There are two inversion formulas from which we can derive an estimate of G.

ii) Part of the thin slice data can be treated as planar data.

iii) The support of G, if it is finit can be determined from the data with high probability.

CHAPTER I

GENERAL THEORY

1.1. Description of the stereological problem.

Spherical particles are dispersed randomly in a three-dimensional body. The classical assumption is that the centers of the spheres are distributed according to a dilute Poisson process. The radius θ of such spheres is distributed according to G independent of everything else. A random probe (line, plane, or thin slice) is cut through the volume. The observed y's are the following:

i) For the random line, its intersection wich a sphere is a line segment, y is half the length of that segment.

ii) For the random plane, y is the radius of the circle of intersection.

iii) For the th... slice, y is the profile radius, that is, the maximum of the radii of the circles of intersection. The problem is to estimate the distribution G from observations $y_1, y_2, \ldots, y_n$.

Stoyan (1979) calculated that the Poisson process has to be very dilute in order for the probability of the spheres to overlap to be small. Thus the classical Poisson model was rather unrealistic. Recently, Mecke and Stoyan (1980) showed that all the formulas derived under the classical assumption of dilute Poisson process also hold under more general and realistic conditions. Their model is the following: let $\{(x_i, \theta_i)\}$ be a marked point process where $\{x_i\}$ is any stationary point process (the centers) and $\{\theta_i\}$ are marks (the radii). This model allows dependencies between diameters and m..dpoint distances, situations that usually occur in nature.

1.2. The classical procedures.

The classical procedures make use of the inversion formula that expresses the particle size distribution G as a function of the distribution H or density h of the observed data. An estimate $\tilde{G}$ is obtained by replacing $H(h)$ in the formula by an estimate $\tilde{H}\ (\tilde{h})$.

1.2a. Random line.

The inversion formula is

$$1 - G(\theta-) = \frac{1}{\theta}\frac{h(\theta)}{h'(0)} \tag{1.2.1}$$

$$\text{where } G(\theta-) = \lim_{t \uparrow \theta} G(t) \ .$$

So, an estimate is

$$1 - \tilde{G}(\theta-) = \frac{1}{\theta}\frac{\tilde{h}(\theta)}{\tilde{h}'(0)} \tag{1.2.2}$$

where $\tilde{h}$ is the spectral derivative of H_n, the sample c.d.f. based on $y_1,\ldots,y_n$.

Note: Spectral differentiation is a stable numerical differentiation procedure developed in a time series context (Anderssen and Bloomfield, 1973). For further discussion, see page

1.2b. Random plane.

The inversion formula is

$$1 - G(\theta) = \frac{\int_\theta^\infty \frac{1}{\sqrt{y^2-\theta^2}}\, dH(y)}{\int_0^\infty \frac{1}{y}\, dH(y)} \ . \tag{1.2.3}$$

So, an estimate is

$$1 - \tilde{G}(\theta) = \frac{\int_\theta^\infty \frac{1}{\sqrt{y^2-\theta^2}}\, d\tilde{H}(y)}{\int_0^\infty \frac{1}{y}\, d\tilde{H}(y)} . \tag{1.2.4}$$

Back substitution: $\tilde{H} = H_n$.

Anderssen and Jakeman: $\tilde{H} = H_n$ smoothed by using a localized Lagrange interpolation.

1.2c. Thin slice of thickness 2τ.

If G is continuous, the inversion formula is

$$1 - G(\theta) = \frac{\int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right) dH(y)}{\int_0^\infty f\left(\frac{y\sqrt{2\pi}}{2\tau}\right) dH(y)} , \tag{1.2.5}$$

where $f(w) = \sqrt{2\pi}\, e^{w^2/2}(1-\Phi(w))$ and Φ is the standard normal cumulative distribution function. So, an estimate is

$$1 - \tilde{G}(\theta) = \frac{\int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right) d\tilde{H}(y)}{\int_0^\infty f\left(\frac{y\sqrt{2\pi}}{2\tau}\right) d\tilde{H}(y)} , \tag{1.2.6}$$

where $\tilde{H}$ is either H_n or a smoothed version of H_n.

For further details, see section 2.2 (random line), section 3.2 (random plane) and section 4.2 (thin slice).

1.3. The basic problem with the classical procedures.

We must remember how H comes about. H is not just any arbitrary distribution but it is the distribution of the data observed through the random probe. In a way that depends on the sampling mechanism (random line, plane or thin slice), H is related to the particle size distribution G. Hence $H = H_G$. In other words, H has to belong to the admissible range $\{H_G\}$ and this admissible range depends on the sampling mechanism.

The basic problem with the classical procedures is that they do not take the structure of the problem into account. They estimate H by the sample c.d.f. or a smoothed version of it and then substitute this estimate $\tilde{H}$ into the inversion formula. This leads to the following:

i) $\tilde{H} \notin \{H_G\}$ with the result that $\tilde{G}$ is not a distribution function. For example, in the random plane case, the back substitution procedure uses $\tilde{H} = H_n$, the sample c.d.f. (1.2.4) becomes

$$1 - G_n(\theta) = \frac{\sum_{y_i \geq \theta} \frac{1}{\sqrt{y_i^2 - \theta^2}}}{\sum_i \frac{1}{y_i}} .$$

As an estimate of G, G_n has the following undesirable properties:

1) $1-G_n$ is increasing in each of the intervals $[y_i, y_{i+1})$ where we assume the y's are already ordered.

2) $1-G_n$ has singularities.

Arguing backward, since G_n is not a distribution function, $H_n \notin \{H_G\}$, and this is the basic problem of back substitution. There is no reason to believe that if we smooth H_n to $\tilde{H}_n$, $\tilde{H}_n \in \{H_G\}$. Thus Anderssen and Jakeman's procedure has the same basic problem (see section 3.3).

ii) The same estimate of H is used regardless of the sampling mechanism and the kind of data (linear, planar or thin slice).

iii) They cannot incorporate additional information concerning G. Such additional information is available in the thin slice case, where if G is discrete with finite support, the support can be determined from the data with high probability (see section 4.5).

1.4. Formulation of the problem of estimating mixing distributions.

Let

$$y_1, y_2, \ldots, y_n \text{ i.i.d.} \sim F_Q$$

where

$$F_Q(y) = \int F_\theta(y)\, dQ(\theta) \;.$$

$Q(\theta)$ is called the mixing distribution. $\{F_\theta\}$ is a parametric family of distributions.

If F_θ has density f_θ, then F_Q has a density $f_Q(y) = \int f_\theta(y) dQ(\theta)$ but it is possible that F_Q has density even though F_θ has no density.

The problem is to estimate $Q(\theta)$ from $y_1, \ldots, y_n$.

For the estimation problem to make sense, the identifiability condition has to be satisfied, i.e.

$$F_{Q_1} = F_{Q_2} \Rightarrow Q_1 = Q_2 .$$

1.5. The connection between the stereological problem and estimation of mixing distributions.

Let $G(\theta)$ be the distribution of the radii. The arrow "$\uparrow$" denotes the event that a sphere is hit by the random probe. Then

$H(y|\theta,\uparrow)$ is the conditional distribution of y given that a sphere of radius θ is hit by the probe.

$P(\uparrow|\theta)$ is the probability that a sphere of radius θ is hit by the probe.

$H(y|\uparrow)$ is the distribution of the observed y's. The arrow "$\uparrow$" is there since y is measured only when a sphere is hit by the probe.

Thus

$$H(y|\uparrow) = \int H(y|\theta,\uparrow)\, dG(\theta|\uparrow) \tag{1.5.1}$$

but

$$dG(\theta|\uparrow) = \frac{P(\uparrow|\theta)dG(\theta)}{\int P(\uparrow|\theta)dG(\theta)} \tag{1.5.2}$$

so

$$H(y|\uparrow) = \frac{\int H(y|\theta,\uparrow)\, P(\uparrow|\theta)\, dG(\theta)}{\int P(\uparrow|\theta)\, dG(\theta)} . \tag{1.5.3}$$

As we will see shortly, in all three cases (random line, plane and thin slice),

$H(y|\theta,\uparrow)$ is a parametric family

$P(\uparrow|\theta)$ is also known.

Therefore (1.5.1) tells us that $G(\theta|\uparrow)$ is a mixing distribution (in the notation of section 1.4, $Q(\theta) = G(\theta|\uparrow)$, $F_\theta(y) = H(y|\theta,\uparrow)$). Then (1.5.2) tells us that estimating $G(\theta|\uparrow)$ is equivalent to estimating $G(\theta)$. We note (1.5.3) is the formula on which stereologists have been focusing their attention, where $P(\uparrow|\theta)$ always cancels with a term in the denominator of $H(y|\theta,\uparrow)$, and the resulting integral equation can be inverted to give $G(\cdot)$ in terms of $H(\cdot|\uparrow)$ or $h(\cdot|\uparrow)$. Call that relationship (1.5.4) for ease of reference. Because of the possibility of cancellation and inversion, no attention is paid to (1.5.1). Consequently, stereologists have for years overlooked the fact that $G(\theta|\uparrow)$ is a mixing distribution. Perhaps this is understandable since it seems unlikely one would not cancel when cancellation is possible. Instead of using only (1.5.4) (see section 1.2), our proposal is that (1.5.1) should also be used along with (1.5.4) since they complement one another.

We now look at the three cases more closely, writing out (1.5.1), (1.5.2), (1.5.3), (1.5.4) explicitly. Their derivations are elementary, and the interested reader is referred to Coleman (1979).

1.5a. Random line.

$$H(y|\theta,\uparrow) \text{ has density } h(y|\theta,\uparrow) = \begin{cases} \dfrac{2y}{\theta^2} & 0 < y \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$P(\uparrow|\theta) \propto \theta^2 .$$

So, we have

$$(1.5.1a) \qquad h(y|\uparrow) = \int_y^\infty \frac{2y}{\theta^2} \, dG(\theta|\uparrow)$$

$$(1.5.2a) \qquad dG(\theta|\uparrow) = \frac{\theta^2 \, dG(\theta)}{\mu_{2G}}$$

where $\mu_{2G} = \int \theta^2 \, dG(\theta)$.

$$(1.5.3a) \qquad h(y|\uparrow) = \int_y^\infty 2y \frac{1}{\mu_{2G}} \, dG(\theta) \; .$$

Notice the cancellation of θ^2.

(1.5.3a) can be inverted to

$$(1.5.4a) \qquad 1 - G(\theta-) = \frac{1}{\theta} \frac{h(\theta|\uparrow)}{h'(0|\uparrow)} \; .$$

1.5b. Random plane.

$H(y|\theta,\uparrow)$ has density $h(y|\theta,\uparrow) = \begin{cases} \dfrac{y}{\theta\sqrt{\theta^2-y^2}} & 0 < y < \theta \\ 0 & \text{otherwise} \end{cases}$

$P(\uparrow|\theta) \propto \theta$.

So we have,

$$(1.5.1b) \qquad h(y|\uparrow) = \int_y^\infty \frac{y}{\theta\sqrt{\theta^2-y^2}} \, dG(\theta|\uparrow) \; ,$$

$$(1.5.2b) \qquad dG(\theta|\uparrow) = \frac{\theta \, dG(\theta)}{\mu_G} \; ,$$

where $\mu_G = \int \theta \, dG(\theta)$.

$$(1.5.3b) \qquad h(y|\uparrow) = \int_y^{\infty} \frac{y}{\sqrt{\theta^2-y^2}} \frac{1}{\mu_G} dG(\theta) \; .$$

Notice the cancellation of θ.

(1.5.3b) can be inverted to

$$(1.5.4b) \qquad 1 - G(\theta) = \int_{\theta}^{\infty} \frac{1}{\sqrt{y^2-\theta^2}} dH(y|\uparrow) \Big/ \int_0^{\infty} \frac{1}{y} dH(y|\uparrow) \; .$$

1.5c. Thin slice of thickness 2τ.

$$1 - H(y|\theta,\uparrow) = \begin{cases} \dfrac{\sqrt{\theta^2-y^2}+\tau}{\theta+\tau} & 0 \le y < \theta \\ 0 & y \ge \theta \; . \end{cases}$$

Note: $H(y|\theta,\uparrow)$ has a jump at θ which can be explained physically. One consequence is that the formulas derived in the literature are valid only when G is continuous (see section 4.2).

$$P(\uparrow|\theta) \propto \theta + \tau \; .$$

So, we have

(1.5.1c) $$1 - H(y|\uparrow) = \int_y^\infty \frac{\sqrt{\theta^2-y^2}+\tau}{\theta+\tau}\, dG(\theta|\uparrow)$$

(1.5.2c) $$dG(\theta|\uparrow) = \frac{(\theta+\tau)\ dG(\theta)}{\mu_G+\tau}$$

(1.5.3c) $$1 - H(y|\uparrow) = \int_y^\infty \frac{\sqrt{\theta^2-y^2}+\tau}{\mu_G+\tau}\, dG(\theta)\ .$$

Notice the cancellation of $\theta+\tau$.

If G is continuous, (1.5.3c) can be inverted to

(1.5.4c) $$1 - G(\theta) = \frac{\int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right) dH(y|\uparrow)}{\int_0^\infty f\left(\frac{y\sqrt{2\pi}}{2}\right) dH(y|\uparrow)}$$

where $f(w) = \sqrt{2\pi}\ e^{w^2/2}\ (1-\Phi(w))$ and Φ is the standard normal distribution function.

If G is mixed or discrete, the correct formulas involve a decomposition of H into its discrete and continuous component (see section 4.2). This makes the estimation problem more complicated but also more interesting, especially in the discrete case. The discrete case is interesting because of the following:

i) There are two inversion formulas from which we can derive an estimate of G.

ii) Part of the thin slice data can be treated as planar data.

iii) The support of G, if it is finite, can be determined from the data with high probability.

For details, see section 4.5.

Note: The case with truncation (i.e. only $y \geq y_0$ is observed) can be reduced to the case without truncation (see sections 2.5, 3.8).

1.6. Summary of notations.

a. For mixture of distributions.

$Q(\theta)$ denotes the mixing distribution.

$\{F_\theta(\cdot)\}$ is the parametric family of distributions that we are mixing.

$\{f_\theta(\cdot)\}$ are the densities of $\{F_\theta(\cdot)\}$ when they exist.

(1.6.1)
$$
\begin{aligned}
F_Q(y) &= \int F_\theta(y)\, dQ(\theta) \\
f_Q(y) &= \int f_\theta(y)\, dQ(\theta) \quad \text{if } f_\theta \text{ exists} \\
&= \frac{d}{dy} \int F_\theta(y)\, dQ(\theta) \quad \text{in general.}
\end{aligned}
$$

b. For the stereological problem.

$G(\theta)$ is the distribution of the radii of the spheres.

$G(\theta|\uparrow)$ - distribution of θ given the sphere is hit by the probe.

$H(y|\theta,\uparrow)$ - distribution of the observed y given that a sphere of radius θ is hit by the random probe.

$H(y|\uparrow)$ - distribution of the observed y's. "$\uparrow$" is there since y is observed only when the probe hits a sphere.

$P(\uparrow|\theta)$ - probability that a sphere of radius θ is hit by the random probe.

Again, small letters g, h denotes densities when they exist.

c. <u>In connecting the two problems</u>, the following identification is employed.

Since $H(y|\uparrow) = \int H(y|\theta,\uparrow)\, dG(\theta|\uparrow)$. By comparing with (1.6.1)

Stereological Problem		Mixture of distributions
$G(\theta\|\uparrow)$	$\longleftrightarrow$	$Q(\theta)$
$H(y\|\theta,\uparrow)$	$\longleftrightarrow$	$F_\theta(y)$
$H(y\|\uparrow)$	$\longleftrightarrow$	$F_Q(y)$

Similar identification for densities

$h(y\|\theta,\uparrow)$	$\longleftrightarrow$	$f_\theta(y)$
$h(y\|\uparrow)$	$\longleftrightarrow$	$f_Q(y)$

1.7. <u>The new approach</u>.

Recall the following: (1.5.1) tells us that $G(\theta|\uparrow)$ is a mixing distribution. (1.5.2) tells us that estimating $G(\theta|\uparrow)$ is equivalent to estimating $G(\theta)$. These motivate the following procedure:

Step 1. Estimate the mixing distribution $G(\theta|\uparrow)$, call that estimate $\tilde{G}(\theta|\uparrow)$.

Step 2. Obtain $\tilde{G}(\theta)$ from $\tilde{G}(\theta|\uparrow)$ using (1.5.2).

Sometimes we obtain $\tilde{H}(y|\uparrow) = F_{\tilde{Q}}(y)$ more directly than $\tilde{G}(\theta|\uparrow) = \tilde{Q}(\theta)$. In that case, we propose an alternate procedure which illustrates how (1.5.1) and (1.5.4) complement one another.

Step 1. Obtain $\tilde{H}(y|\uparrow) = F_{\tilde{Q}}(y)$.

Step 2. Substitute $\tilde{H}$ into the inversion formula (1.5.4) to obtain $\tilde{G}$.

This procedure works nicely in the random line case (see section 2.3).

1.8. Nonparametric estimation of mixing distributions.

The purpose of this section is to describe some nonparametric procedures which we will apply later (either directly or with slight modification).

For notations, see section 1.6.

1.8a. Nonparametric maximum likelihood estimate (MLE).

Assuming F_Q has density f_θ. Then F_Q has density f_Q, $f_Q(y) = \int f_\theta(y)\, dQ(\theta)$. We have $y_1, y_2, \ldots, y_n$, i.i.d. $\sim f_Q$. The procedure is to find $\hat{f}$ that maximizes $\Pi_{i=1}^n f(y_i)$ subject to $f \in \{f_Q\}$

We now remark:

i) Maximizing $\Pi_{i=1}^n f(y_i)$ determines $\hat{f}$ only up to $\hat{f}(y_1), \ldots, \hat{f}(y_n)$, further argument is needed in each special case to show that they determine $\hat{f}$ uniquely.

ii) By identifiability, $\hat{f} = f_{\hat{Q}}$ determines $\hat{Q}$, but this is only an existence theorem, so we need a procedure for finding $\hat{Q}$ from $\hat{f}$. This may not be easy in general, but for the problem of unfolding particle size distribution, something nice happens. Recall that (1.5.1) tells us we have a problem in estimating a mixing distribution by the identification $F_\theta(y) = H(y|\theta, \uparrow)$ and $Q(\theta) = G(\theta|\uparrow)$, $F_Q(y) = H(y|\uparrow)$ but, after all, it is $G(\theta)$, not $G(\theta|\uparrow)$, that we are after, and (1.5.4) gives $G(\cdot)$ in terms of $H(\cdot|\uparrow)$.

The above illustrates how (1.5.1) and (1.5.4) complement one another. (1.5.1) provides the point of view that the distribution of y is a mixture which we can estimate using existing methods. Then (1.5.4) performs the inversion for us to give an estimate of G.

iii) For finite mixture with known support points, we can use the EM algorithm (Dempster, Laird and Rubin, 1977) to find $\hat{Q}$ directly without the intermediate step $\hat{f}$. This is useful since frequently, a mixture problem can be reduced to a finite mixture with known support points.

We now give a brief description of the EM algorithm in its general setting. Let

$$(\underset{\sim}{y},\underset{\sim}{z}) \sim f_\phi(\underset{\sim}{y},\underset{\sim}{z})$$

$$\ell_\phi(\underset{\sim}{y},\underset{\sim}{z}) = \log f_\phi(\underset{\sim}{y},\underset{\sim}{z})$$

where $\underset{\sim}{y}$ are the observed data, $(\underset{\sim}{y},\underset{\sim}{z})$ are the complete data. Define

$$Q(\phi'|\phi) = E_\phi(\ell_{\phi'}(\underset{\sim}{y},\underset{\sim}{z})|\underset{\sim}{y}) \ .$$

Each iteration $\phi^{(m)} \to \phi^{(m+1)}$ of the EM algorithm involves two steps which we call the expectation step (E-step) and the maximization step (M-step).

E-step: Compute $Q(\phi|\phi^{(m)})$.

M-step: Choose $\phi^{(m+1)}$ that maximizes $Q(\phi|\phi^{(m)})$.

The heuristic idea here is that we would like to choose the ϕ^* that maximizes $\ell_\phi(\underset{\sim}{y},\underset{\sim}{z})$. Since we do not know $\ell_\phi(\underset{\sim}{y},\underset{\sim}{z})$, we maximize its current expectation instead given the data $\underset{\sim}{y}$ and the current fit $\phi^{(m)}$.

When we specialize to the case of a finite mixture,

$$\phi = (\theta_1,\theta_2,\ldots,\theta_k,P_1,\ldots,P_{k-1})$$

where $\theta_1,\ldots,\theta_k$ can either be known or unknown,

$$0 < P_i < 1$$

$$P_1 + \cdots + P_{k-1} < 1$$

$$P_k = 1-(P_1 + \cdots + P_{k-1}) \ .$$

Let Z_i, $i = 1,\ldots,n$ be i.i.d. $P_\phi(z_i=j) = P_j$ and $y_i | z_i = j \sim f_{\phi_j}(y_i)$, $i = 1,\ldots,n$ independent. Then

$$f_\phi(\underset{\sim}{y},\underset{\sim}{z}) = \sum_{j=1}^{k} P_j^{n_j} \prod_{\substack{i \\ z_i=j}} f_{\theta_j}(y_i) \text{ where } n_j = \#\, i \text{ such that } Z_i = j \ ,$$

and

$$\ell_\phi(\underset{\sim}{y},\underset{\sim}{z}) = \underbrace{\sum_{j=1}^{k} n_j \log P_j}_{\substack{\text{depends on } \phi \\ \text{only through} \\ P_1,\ldots,P_{k-1}}} + \underbrace{\sum_{j=1}^{k} \sum_{\substack{i \\ z_i=j}} \log f_{\theta_j}(y_i)}_{\substack{\text{depends on } \phi \\ \text{only through} \\ \theta_1,\ldots,\theta_k}} \ .$$

Thus to apply the EM algorithm, we maximize the conditional expectation of the two sums separately given $\underset{\sim}{y}$ and $\phi^{(m)}$. In particular,

$$P_j^{(m+1)} = \frac{E_{\phi^{(m)}}(n_j | \underset{\sim}{y})}{n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{P_j^{(m)} f_{\theta_j^{(m)}}(y_i)}{\sum_{\ell=1}^{k} P_\ell^{(m)} f_{\theta_\ell^{(m)}}(y_i)} \ .$$

If $\theta_1,\ldots,\theta_k$ are known, it reduces to

$$P_j^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} \frac{P_j^{(m)} f_{\theta_j}(y_i)}{\sum_{\ell=1}^{k} P_\ell^{(m)} f_{\theta_\ell}(y_i)} \ .$$

b) Minimum distance method.

Motivation: Since $y_1, y_2, \ldots, y_n$ i.i.d. $\sim F_Q$ and F_Q determines Q by identifiability, all we need to do is to obtain an estimate $\hat{F}$ of F_Q from $y_1, y_2, \ldots, y_n$. However, we must remember that $\hat{F}$ has to belong to the admissible range $\{F_Q\}$. Thus, for example, if F_θ is continuous, then so is $F_Q = \int F_\theta \, dQ(\theta)$ regardless of Q. So, in this case, the sample c.d.f. $F_n \notin \{F_Q\}$. F_n is the nonparametric MLE of F, but if $\hat{Q}$ is the nonparametric MLE of Q, then the corresponding $F_{\hat{Q}}$ would be the restricted MLE of F subject to $F \in \{F_Q\}$ and not F_n, the unrestricted MLE. Since the restricted MLE may be hard to find and sometimes can be inconsistent, the method of minimum distance is proposed.

The method can be described as follows: Among all Q that put positive probability at n points, find $\hat{Q}_n$ that minimizes $d(F_Q, F_n)$ where $d(\cdot,\cdot)$ is some distance function and F_n is the sample c.d.f. of the y's. There are variations due to the choice of $d(\cdot,\cdot)$ and as to whether the n points of support are pre-specified or are among the parameters in the minimization problem.

i) Deely and Kruse's approach.

$d(F, F_n) = \|F - F_n\|$, the sup norm. The n points of support $\theta_{1n}, \theta_{2n}, \ldots, \theta_{nn}$ are pre-specified. They are so chosen that if $\mathcal{Q}_n = \{$discrete d.f. putting mass at $\theta_{1n}, \ldots, \theta_{nn}\}$ then there exists a sequence $\{Q_n\}$, $Q_n \in \mathcal{Q}_n$ and $Q_n \xrightarrow{\mathcal{D}} Q$.

ii) Choi and Bulgren's approach.

$d(F,F_n) = \int (F(y)-F_n(y))^2 \, dF_n(y)$, the Wolfowitz distance. The n points of support $\theta_{1n},\ldots,\theta_{nn}$ are among the parameters.

Under the regularity conditions,

1) identifiability.

2) $\lim_{\theta \to \infty} F_\theta(y) = F_\infty(y)$ exists, $F_\infty(\cdot)$ is not a distribution.

3) $F_\theta(y)$ continuous in (θ,y)

both approaches give a consistent estimate of Q.

1.9. The advantages of the new approach over the classical approach.

Our new approach takes the structure of the problem into account. Matching the drawbacks listed in section 1.3 of the classical procedures, we have the following:

i) $\tilde{H} \in \{H_G\} = F_Q$. The importance of having $\tilde{H}$ lie within the admissible range is highlighted in the random plane case where we want to bootstrap the distribution of a stereological estimate. In section 3.7, we show that the usual way of resampling from the sample c.d.f. in the bootstrap does not work. We find out that it is crucial to resample from $\tilde{H}$ belonging to the admissible range.

ii) For different kinds of data (linear, planar or thin slice), H is the mixture of different parametric family of distributions, hence the estimate $\tilde{H}$ would also be different.

iii) The new approach can easily incorporate additional information concerning G to give a better estimate. For example, if we know the support of G, hence also of $Q = G(\cdot|\uparrow)$, is $\{\theta_1,\ldots,\theta_m\}$, then the MLE

$\hat{Q}$ maximizes $\Sigma_{i=1}^{n} f_Q(y_i)$ among all Q that have support $\{\theta_1,\ldots,\theta_m\}$, (instead of $\{\theta_{1n},\ldots,\theta_{nn}\}$). Similarly, the minimum distance estimate $\hat{Q}$ minimizes $d(F_Q,F_n)$ among all Q that have support $\{\theta_1,\ldots,\theta_m\}$. Since m is fixed and does not grow with the sample size n, the reduction in the amount of computation can be substantial.

1.10. Some peculiar features.

It is worth pointing out that there are some peculiar features which make the problem unique and more than just a routine special case of estimating a mixing distribution.

i) The integral equation (1.5.3) can be inverted to (1.5.4) which gives $G(\cdot)$ in terms of $H(\cdot|\dagger)$ ($h(\cdot|\dagger)$ in the random line case). Thus bacr ıbstitution is possible. That is, we can just find an estimate of $H(\cdot|\dagger)$ ($h(\cdot|\dagger)$), call that $\tilde{H}(\cdot|\dagger)(\tilde{h}(\cdot|\dagger))$, plug it into (1.5.4) and it will give us an estimate $\tilde{G}$ of G. In particular, if $\tilde{H} = \hat{H} = F_{\hat{Q}}$, the restricted MLE (recall the identification $Q(\theta) = G(\theta|\dagger)$, $F_\theta(y) = H(y|\theta,\dagger)$, $F_Q(y) = H(y|\dagger)$), then $\tilde{G} = \hat{G}$, the nonparametric MLE of G. A work of caution is that if $\tilde{H} \notin \{F_Q\}$, $\tilde{G}$ will not be a distribution function.

ii) For the random plane case.

$$h(y|\theta,\dagger) = \begin{cases} \dfrac{y}{\theta\sqrt{\theta^2-y^2}} & 0 \le y < \theta \\ \\ 0 & \text{otherwise} \end{cases},$$

has a singularity at $y = \theta$.

iii) For the thin slice case.

$H(y|\theta,\dagger)$ has a jump at $y = \theta$.

iv) As a problem in estimating a mixing distribution, the $F_\theta(y)$ i.e. the $H(y|\theta,\dagger)$ in the three cases are so peculiar that no one would think of doing these special cases if they did not arise out of a practical application, which is the case here.

CHAPTER II

THE RANDOM LINE CASE

2.1. The basic formulas.

(2.1.1)
$$h_Y(y|\uparrow) = \int_y^\infty \frac{2y}{\theta^2}\, dG(\theta|\uparrow)$$

(2.1.2)
$$dG(\theta|\uparrow) = \frac{\theta^2\, dG(\theta)}{\mu_{2G}}$$

where $\mu_{2G} = \int \theta^2\, dG(\theta)$.

(2.1.3)
$$h_Y(y|\uparrow) = \int_y^\infty \frac{2y}{\mu_{2G}}\, dG(\theta) \ .$$

Equation (2.1.3) can be inverted to give

$$1 - G(\theta-) = \frac{h_Y(\theta|\uparrow)}{\theta}\ \frac{\mu_{2G}}{2} \ .$$

Taking the limit as $\theta \to 0$ gives $1 = h'_Y(0|\uparrow)\ \frac{\mu_{2G}}{2}$. So

(2.1.4)
$$1 - G(\theta-) = \frac{1}{\theta}\ \frac{h_Y(\theta|\uparrow)}{h'_Y(0|\uparrow)} \ .$$

2.2. Survey of the Literature.

Since $y_1, y_2, \ldots, y_n$ i.i.d. $\sim h_Y(y|\dagger)$, we can obtain some estimate $\tilde{h}_Y(\cdot|\dagger)$ from the data. Then an estimate of G is

$$1 - \tilde{G}(\theta-) = \frac{1}{\theta} \frac{\tilde{h}_Y(\theta|\dagger)}{\tilde{h}_Y'(0|\dagger)} . \tag{2.2.1}$$

This procedure is found to be unsatisfactory. Some objections are the following:

i) The choice of $\tilde{h}_Y(\cdot|\dagger)$ is arbitrary.

ii) As far as we know, no stereologist attempts to find the whole density $\tilde{h}_Y(\cdot|\dagger)$. They just differentiate $H_n(y)$ (the sample c.d.f. of the y's) numerically to obtain $\tilde{h}(\cdot|\dagger)$ on an even grid. The numerical procedure proposed is spectral differentiation: a stable procedure developed in a time series context. (Anderssen and Bloomfield, 1973).

iii) Since $\tilde{h}(\cdot|\dagger)$ is evaluated only on an even grid, we also need to estimate $h_Y'(0|\dagger)$ which is very hard to estimate from the data.

iv) If $G(\theta)$ has density $g(\theta)$, then from (2.1.4),

$$g(\theta) = \frac{-1}{h_Y'(0|\dagger)} \frac{d}{d\theta} \left(\frac{h_Y(\theta|\dagger)}{\theta}\right) = \frac{-1}{h_Y'(0|\dagger)} \frac{d}{d\theta} \left(\frac{\frac{d}{d\theta} H_Y(\theta|\dagger)}{\theta}\right) .$$

Thus to estimate $g(\theta)$ starting from the sample c.d.f. H_n, we need to obtain the spectral derivative of order 2. Simulation results of Jakeman and Anderssen (1976) for a bimodal $g(\theta)$ shows that the procedure is unsatisfactory.

v) Another factor which contributes to the failure is the effect of the factor $\frac{1}{\theta}$ in the neighborhood of 0.

Because of the above, many writers (e.g. Nicholson, 1970, Watson, 1971, Moran, 1971, Jakeman and Anderssen, 1976) have advised that linear probe data should be avoided.

2.3. Nonparametric MLE.

2.3a. Derivation and computation.

The nonparametric MLE of G can be derived as follows:

i) Make the transformation $Z = Y^2$.

Since
$$h_Y(y|\theta,\dagger) = \frac{2y}{\theta^2} \qquad 0 \le y \le \theta$$
$$h_Z(z|\theta,\dagger) = \frac{1}{\theta^2} \qquad 0 \le z \le \theta^2 .$$

So
$$h_Z(z|\dagger) = \int_{\sqrt{z}}^{\infty} \frac{1}{\theta^2}\, dG(\theta|\dagger) .$$

ii) Let $\xi = \theta^2$.

Then
$$h_Z(z|\dagger) = \int_{z}^{\infty} \frac{1}{\xi}\, dG(\xi|\dagger)$$

which tells us that $h_Z(\cdot|\dagger)$ is a mixture of uniforms.

iii) Densities which are mixtures of uniforms coincide with left continuous nonincreasing densities. The nonparametric MLE in this case is well known (Barlow, Bartholomew, Bremner and Brunk, 1972, P. 223-228).

Fact 1: $\hat{H}_Z(\cdot|\dagger)$ is the least concave majorant of, $H_n(\cdot)$, the empirical c.d.f. which put mass $\frac{1}{n}$ at $z = y_1^2, y_2^2, \ldots, y_n^2$.

Fact 2: $\hat{h}_Z(\cdot|\dagger)$ is a left continuous nonincreasing step function with jumps at $z = y_1^2, y_2^2, \ldots, y_n^2$. In fact, $\hat{h}_Z(z|\dagger)$ is the slope of $\hat{H}_Z(\cdot|\dagger)$ at $z-$.

iv) An algorithm for computing the MLE, like the "Up-and-Down Blocks" algorithm can be found in Barlow, Bartholomew, Bremner and Brunk (P. 72). This algorithm is easily programmed.

v) By invariance, the nonparametric MLE of $h_Y(\cdot|\uparrow)$ is

$$\hat{h}_Y(y|\uparrow) = 2y\hat{h}_Z(y^2|\uparrow) \tag{2.3.1}$$

$$\text{Fact 2} \Rightarrow \hat{h}'_Y(0|\uparrow) = 2\hat{h}_Z(y_1^2|\uparrow) \tag{2.3.2}$$

where we assume $y_1, y_2, \ldots, y_n$ is already arranged in increasing order.

Substituting (2.3.1) and (2.3.2) and (2.2.1) gives

$$1 - \hat{G}(\theta-) = \frac{2\theta\hat{h}_Z(\theta^2|\uparrow)}{\theta\, 2\hat{h}_Z(y_1^2|\uparrow)} \Rightarrow$$

$$1 - \hat{G}(\theta-) = \frac{\hat{h}_Z(\theta^2|\uparrow)}{\hat{h}_Z(y_1^2|\uparrow)} . \tag{2.3.3}$$

Fact 2 combined with (2.3.3) tells us

Fact 3: $\hat{G}$ is discrete, putting mass

$$\frac{\hat{h}_Z(y_1^2|\uparrow) - \hat{h}_Z(y_2^2\uparrow|)}{\hat{h}_Z(y_1^2|\uparrow)}, \ldots,$$

$$\frac{\hat{h}_Z(y_{n-1}^2|\uparrow) - \hat{h}_Z(y_n^2|\uparrow)}{\hat{h}_Z(y_1^2|\uparrow)}, \quad \frac{\hat{h}_Z(y_n^2|\uparrow)}{\hat{h}_Z(y_1^2|\uparrow)}$$

at the points $y_1, \ldots, y_{n-1}, y_n$.

2.3b. Argument for using MLE.

The nonparametric MLE seems to provide an answer to the objections raised against the classical procedures (section 2.2). Objections i, ii, iii, are answered since we choose $\tilde{h}_Y(\cdot|\uparrow) = \hat{h}(\cdot|\uparrow)$, the nonparametric MLE which is well defined. Objection iv is answered since in order to estimate $g(\theta)$, we differentiate (2.3.3) which requires a spectral derivative of order 1 instead of 2.

Objection v is also answered since the factor $\frac{1}{\theta}$ is cancelled in (2.3.3).

2.3c. Consistency of MLE.

From (2.3.3), consistency of $\hat{G}$ follows from consistency of $\hat{h}_Z(\cdot|\uparrow)$ which is proved in Barlow, Bartholomew, Bremner and Brunk (P. 228).

2.4. Simulation results.

The stationary distribution

$$g(\theta) = \frac{\theta}{c^2} \exp(-\frac{\theta^2}{2c^2})$$

is employed so that $h_Y(\cdot|\uparrow) = g(\cdot)$. In what follows, $c = 4.0$. The sample size n is taken to be 100. Thus $y_1, y_2, \ldots, y_{100}$ are generated according to $h_Y(\cdot|\uparrow)$.

The nonparametric MLE is computed using fact 3 of section 2.3, where $\hat{h}_Z(\cdot|\uparrow)$ is computed using the "Up-and-Down Block" algorithm. Averaging over 100 trials, the result is as follows.

TABLE 1

The average of the MLE $\hat{G}(\theta)$ and the average percentage error over 100 samples of size 100 drawn from the stationary distribution

$$h(\theta) = g(\theta) = \frac{\theta}{16} e^{-\theta^2/32}$$

θ	$G(\theta)$	$\hat{G}(\theta)$	% ERROR
0.6070	0.0114	0.3470	3025.25
1.2139	0.0450	0.4807	1015.94
1.8209	0.0984	0.5214	462.291
2.4279	0.1682	0.5636	252.551
3.0349	0.2501	0.6039	153.999
3.6418	0.3393	0.6481	99.8950
4.2488	0.4311	0.7029	69.4484
4.8558	0.5214	0.7499	47.3913
5.4627	0.6065	0.7900	35.1072
6.0697	0.6838	0.8329	25.4512
6.6767	0.7517	0.8726	18.2788
7.2837	0.8095	0.9030	12.6061
7.8906	0.8571	0.9298	9.2314
8.4976	0.8953	0.9457	6.3706
9.1046	0.9250	0.9602	4.5193
9.7115	0.9475	0.9712	3.3260
10.3185	0.9641	0.9833	2.2494
10.9255	0.9760	0.9899	1.5802
11.5324	0.9843	0.9942	1.1271
12.1394	0.9900	0.9969	0.8149

The results look horrible but this has to do with the fact that the MLE is badly "spiked" at the mode. This peculiar "peaking" near the mode of the MLE was observed by Wegman (1970a,1970b). He also suggested a partial solution by requiring the estimate to have a modal interval of length ε, where ε is some fixed positive number.

By pooling the first ten observations into one block and then applying the "Up-and-Down Block" algorithm, which is essentially a modified version of Wegman's procedure, we are able to obtain much improved estimation results. Call this estimate $\hat{G}_1$. We compare $\hat{G}_1$ with various versions of the classical estimator.

To describe the various versions of the classical estimator, we need the following:

From 0 to 12.1394, the 0.99th quantile of G, we form a grid of 101 equally spaced points $t_1 = 0, t_2, \ldots, t_{100}, t_{101} = 12.1394$. Let H_n be the sample c.d.f. of $H_Y(\cdot|+)$ so that $\{H_n(t_k);\ k = 1, \ldots, 101\}$ represents the values of H_n on that grid. Since we do not have exact data, that is, we do not have $H(t_k)$ but only $H_n(t_k) = H(t_k) + \varepsilon_k$, we need spectral differentiation, a numerical differentiation procedure for non-exact data. Let

$$h^*(t_k) = \left[\frac{dH_n(t)}{dt}\right]^*_{t=t_k}$$

= spectral derivative of order 1 evaluated at $t = t_k$ for the data $\{H_n(t_k)\}$.

Note: In evaluating the spectral derivative, we first take away a least-squares 5^{th} degree polynomial trend $P(\cdot)$ satisfying the four end conditions

$$P(t_1) = H_n(t_1)$$

$$P(t_{101}) = H_n(101)$$

$$P'(t_1) = \frac{H_n(t_2) - H_n(t_1)}{t_2 - t_1}$$

$$P'(t_{101}) = \frac{H_n(t_{101}) - H_n(t_{100})}{t_{101} - t_{100}}\ .$$

By removing an average trend $P(t)$ about which the data tend to vary in a near stationary manner, the residual data $\{v_k\} = \{H_n(t_k)-P(t_k)\}$ has the following structure which is essential for spectral differentiation to be applicable,

$$v_k = u(t_k) + \varepsilon_k \tag{2.4.1}$$

where $u(t)$, the smooth function, is a stationary stochastic process with continuous parameter t and zero expectation and $\{\varepsilon_k\}$, the measurement error, is defined by a stationary stochastic process with discrete parameter k and zero expectation.

This permits us to apply spectral differentiation to $\{v_k\}$ to obtain $u'^*(t_k)$, the spectral derivative of $u(t)$. Since

$$H_n(t_k) = H(t_k) + \varepsilon_k$$

$$v_k = H_n(t_k) - P(t_k)$$

$$= H(t_k) - P(t_k) + \varepsilon_k \ .$$

Compare with (2.4.1), we have

$$H(t_k) = u(t_k) + P(t_k) \ .$$

So $h^*(t_k) = u'^*(t_k) + P'(t_k)$, which we call the spectral derivative of $H(t)$, is an estimate of $h(t_k) = H'(t_k)$.

A natural candidate for the trend is a least-squares polynomial. The first two end conditions we imposed are natural end conditions, while the last two represent a numerical expedient introduced to cope with a shortcoming in spectral differentiation. Since the relative accuracy of a spectral derivative deteriorates only in the neighborhood

of the endpoints, the last two end conditions ensure that the value of the derivative to be estimated by spectral differentiation is small compared with the actual derivative. (That is, $u' = H'-P'$ is small compared with H' at the endpoints). Since, we have four end conditions, we remove a least-squares 5^{th} degree polynomial so that we have two free parameters.

We are now ready to describe the various versions of the classical estimator, $\tilde{G}$.

Version 1. $\tilde{G}_1$. Since G is continuous,

$$1 - G(\theta) = \frac{h_Y(\theta|\uparrow)}{\theta\ h_Y'(0|\uparrow)}$$

$$1 - \tilde{G}_1(t_k) = \frac{h^*(t_k)}{t_k\ h_Y'(0|\uparrow)}\ , \quad k = 1,\ldots,101\ .$$

That is, we use the true value of $h_Y'(0|\uparrow)$ instead of estimating it. As a result, it would give an overly optimistic assessment of the method.

Version 2. $\tilde{G}_2$.

$$1 - \tilde{G}_2(t_k) = \frac{h^*(t_k)}{t_k\left(\dfrac{h^*(t_2) - h^*(t_1)}{t_2 - t_1}\right)}\ , \quad k = 1,\ldots,101\ .$$

That is, $h_Y'(0|\uparrow)$ is estimated by

$$\frac{h^*(t_2) - h^*(t_1)}{t_2 - t_1}\ .$$

Version 3. $\tilde{G}_3$. $h_Y'(0|\uparrow)$ is estimated by

$$\frac{h^*(t_6) - h^*(t_1)}{t_6 - t_1} .$$

Version 4. $\tilde{G}_4$. $h_Y'(0|\uparrow) = H_Y''(0|\uparrow)$ is estimated as follows:

Let

$$\left[\frac{d^2 H_n(t)}{dt^2}\right]^*_{t=t_k} = \text{spectral derivative of order 2 evaluated at } t = t_k \text{ for the data } \{H_n(t_k)\} .$$

Define $i = \min\{k: \left[\frac{d^2 H_n(t)}{dt^2}\right]^*_{t=t_k} > 0\}$. Then $\tilde{h}_Y'(0|\uparrow) = \left[\frac{d^2 H_n(t)}{dt^2}\right]^*_{t=t_i}$.

The same 5^{th} degree polynomial trend described earlier is first removed before we compute the derivative.

Version 5. $\tilde{G}_5$. Same as version 4 except that we first take away a least-squares 7^{th} degree polynomial trend $P(\cdot)$ satisfying the restrictions

$$P(t_1) = H_n(t_1) \qquad P(t_{101}) = H_n(t_{101})$$

$$P'(t_1) = \frac{H_n(t_2) - H_n(t_1)}{t_2 - t_1} \qquad P'(t_{101}) = \frac{H_n(t_{101}) - H_n(t_{100})}{t_{101} - t_{100}}$$

$$P''(t_1) = \frac{h^*(t_6) - h^*(t_1)}{t_6 - t_1} \qquad P''(t_{101}) = \frac{h^*(t_{101}) - h^*(t_{96})}{t_{101} - t_{96}}$$

where we recall h^* is the spectral derivative of order 1.

Version 6. $\tilde{G}_6$. This is a combination of version 3 and version 4. For small values of k, say $k = 6$, $\tilde{G}_4(t_6)$ can sometimes be negative,

corresponding to those cases where $\tilde{h}_Y'(0|+)$ is small. Thus when $\tilde{G}_4(t_6)$ is negative, estimate $h_Y'(0|+)$ by $\dfrac{h^*(t_6)}{t_6}$ instead, forcing $\tilde{G}_6(t_6) = 0$. A comparison of $\hat{G}_1$ with $\tilde{G}_1$–$\tilde{G}_6$ is given in Table 2.

TABLE 2

A comparison of the modified MLE $\hat{G}_1(\theta)$ with six versions of the classical estimate ($\tilde{G}_i(\theta)$, $i=1,\ldots,6$) in terms of the average of the estimates and their average percentage error (shown on even rows) over 100 samples of size 100.

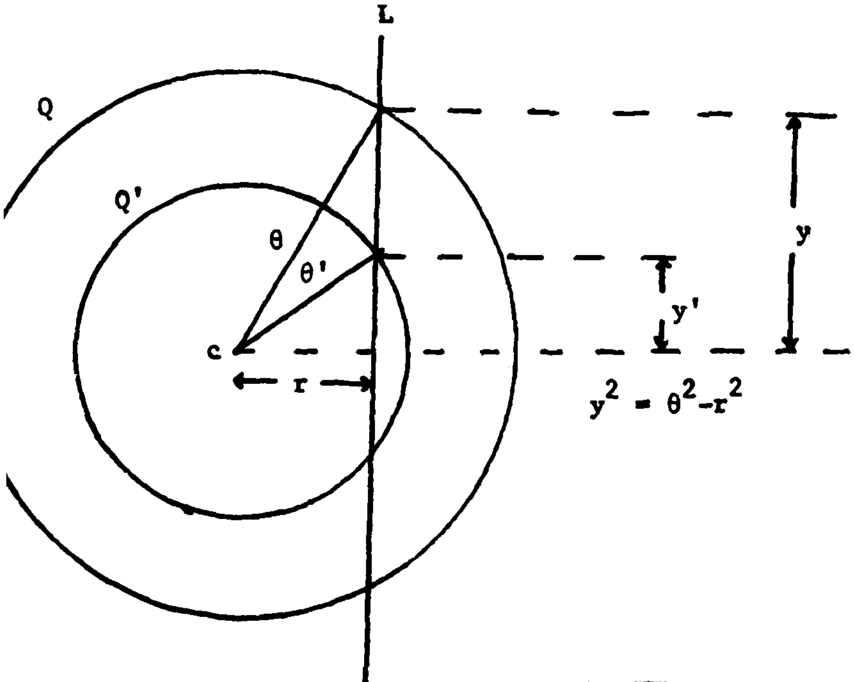
θ	$G(\theta)$	$\hat{G}_1(\theta)$	$\tilde{G}_1(\theta)$	$\tilde{G}_2(\theta)$	$\tilde{G}_3(\theta)$	$\tilde{G}_4(\theta)$	$\tilde{G}_5(\theta)$	$\tilde{G}_6(\theta)$
1.214	0.045	0.014	−0.024	0.166	0.200	−2.332	0.149	0.247
		126.22	357.79	458.92	458.46	5831.0	439.06	495.22
2.428	0.168	0.185	0.192	0.291	0.342	−1.288	0.310	0.379
		96.291	44.720	159.95	145.88	1055.1	131.31	156.07
3.642	0.339	0.356	0.396	0.433	0.486	−0.665	0.464	0.515
		44.985	17.923	79.617	69.811	385.96	63.998	72.812
4.856	0.521	0.545	0.556	0.566	0.615	−0.381	0.597	0.635
		24.813	7.686	41.525	34.796	215.05	32.154	35.814
6.070	0.684	0.701	0.678	0.676	0.716	−0.108	0.701	0.731
		14.602	4.248	23.312	18.211	134.08	17.553	18.274
7.284	0.810	0.823	0.795	0.789	0.818	0.250	0.807	0.827
		6.322	3.119	13.045	9.930	77.735	9.811	9.737
8.498	0.895	0.899	0.885	0.883	0.898	0.554	0.892	0.904
		4.012	1.959	6.432	4.853	42.178	4.889	4.750
9.712	0.948	0.948	0.952	0.953	0.959	0.329	0.957	0.961
		2.657	1.740	2.972	2.556	15.765	2.490	2.549
10.926	0.976	0.981	0.981	0.987	0.986	0.948	0.985	0.987
		1.361	1.257	1.366	1.337	5.403	1.321	1.325
12.139	0.990	0.994	0.997	0.996	0.996	1.015	0.997	0.996
		0.794	1.735	1.649	0.650	3.573	1.677	1.613

<u>Conclusion.</u> Except for $\tilde{G}_1$ which we cannot use in practice since $h'_Y(0|+)$ is unknown, the nonparametric MLE $\hat{G}$ performs better than the classical estimators in terms of both bias and percentage error.

Another advantage of the MLE is that it actually gives an estimate $\hat{G}$ of G whereas the classical procedures estimate G only on an even grid.

2.5. Truncation.

Suppose y is observed only if $y \geq y_0$. Then if $\theta < y_0$, we can never observe y. Whereas if we have a sphere Q of radius $\theta \geq y_0$ with center at C, then we observe $y \geq y_0$ if and only if the linear probe L hits a sphere Q' of radius $\theta' = \sqrt{\theta^2 - y_0^2}$ also centered at C. The intersection of L with Q' is a line segment, let y' be half the length of that segment. The following diagram shows what is going on.



$$y^2 = \theta^2 - r^2 \qquad y' = \sqrt{\theta'^2 - r^2} = \sqrt{\theta^2 - y_0^2 - r^2} = \sqrt{y^2 - y_0^2}$$

Thus if we replace Y by $\sqrt{Y^2 - y_0^2}$ and $G(\theta)$ by the distribution of $\sqrt{\theta^2 - y_0^2}$ given that $\theta \geq y_0$, we will be back to the case with no truncation.

CHAPTER III

THE RANDOM PLANE CASE

3.1. The basic formulas.

$$h(y|\uparrow) = \int_y^\infty \frac{y}{\theta\sqrt{\theta^2-y^2}}\, dG(\theta|\uparrow) \tag{3.1.1}$$

$$dG(\theta|\uparrow) = \frac{\theta\, dG(\theta)}{\mu_G} \tag{3.1.2}$$

where $\mu_G = \int \theta\, dG(\theta)$. So

$$h(y|\uparrow) = \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}}\, \frac{1}{\mu_G}\, dG(\theta) \; . \tag{3.1.3}$$

Again, (3.1.3) can be inverted to give

$$1 - G(\theta) = \frac{2\mu_G}{\pi} \int_\theta^\infty \frac{h(y|\uparrow)}{\sqrt{y^2-\theta^2}}\, dy \; .$$

Setting $\theta = 0$, we have

$$1 = \frac{2\mu_G}{\pi} \int_0^\infty \frac{1}{y}\, h(y|\uparrow) dy \; . \tag{3.1.4}$$

So

$$1 - G(\theta) = \frac{\displaystyle\int_\theta^\infty \frac{h(y|\uparrow)}{\sqrt{y^2-\theta^2}}\, dy}{\displaystyle\int_0^\infty \frac{1}{y}\, h(y|\uparrow) dy} \; . \tag{3.1.5}$$

If G has density g, then

$$(3.1.6) \qquad g(\theta) = \frac{\frac{d}{d\theta}\int_\theta^\infty \frac{h(y|\dagger)}{\sqrt{y^2-\theta^2}}\,dy}{\int_0^\infty \frac{1}{y}\,h(y|\dagger)dy} .$$

3.2. Survey of the literature.

Since $y_1, y_2, \ldots, y_n$ i.i.d. $\sim h(y|\dagger)$, one can replaced $h(y|\dagger)dy$ in (3.1.5) by $dH_n(y)$ where H_n is the sample c.d.f. This is known as back substitution, we have

$$1 - G_n(\theta) = \frac{\sum\limits_{y_i \geq \theta} \frac{1}{\sqrt{y_i^2-\theta^2}}}{\sum\limits_i \frac{1}{y_i}} .$$

$1-G_n$ is unsatisfactory because $1-G$ is decreasing but $1-G_n$ is increasing in each of the intervals $[y_{n,i}, y_{n,i+1})$ where $y_{n,1} < y_{n,2} < \cdots < y_{n,n}$ are the ordered y's. Moreover, $1 - G_n$ has singularities, it goes to infinity as $\theta \uparrow y_i$, $i = 1,\ldots,n$.

Anderssen and Jakeman (1975) explained the poor performance of back substitution by its use of a simple rectangular quadrature approximation rather than a more appropriate approximation which coped with the singularity of the integrand $\frac{1}{\sqrt{y^2-\theta^2}}$ at $y = \theta$. They proposed a method based on product integration. H_n was smoothed to $\tilde{H}_n$ using a localized Lagrange interpolation and $h(y|\dagger)dy$ in (3.1.5) was replaced by $d\tilde{H}_n(y)$. They demonstrated some desirable properties for this class of estimators.

To estimate $g(\theta)$, Anderssen and Jakeman proposed estimating $G(\theta)$ on an even grid by their product integration estimator, before using spectral differentiation.

3.3. Characterization of $h(\cdot|\uparrow)$.

We can estimate G by first estimating $H(\cdot|\uparrow)$ and then using the inversion formula (3.1.5). However, $\{H(\cdot|\uparrow)\}$ is a smaller class than the class of all distribution functions and we have to take that into consideration when we estimate $H(\cdot|\uparrow)$. Thus it is important to have a characterization of $H(\cdot|\uparrow)$.

Since $H(y|\theta,\uparrow)$ has density

$$h(y|\theta,\uparrow) = \frac{y}{\theta\sqrt{\theta^2-y^2}} \qquad 0 < y < \theta \ .$$

$H(y|\uparrow)$ has density

$$h(y|\uparrow) = \int_y^\infty \frac{y}{\theta\sqrt{\theta^2-y^2}}\, dG(\theta|\uparrow) \ .$$

$h(y|\uparrow)$ satisfies the following conditions:

c1) $$\int_0^\infty h(y|\uparrow)\, dy = 1 \ , \ h(y|\uparrow) \geq 0$$

c2) $h(y|\uparrow)$ is right continuous

c3) $$\int_0^\infty \frac{1}{y}\, h(y|\uparrow)\, dy < \infty$$

c4) $$\int_\theta^\infty \frac{h(y|\uparrow)}{\sqrt{y^2-\theta^2}}\, dy \text{ is nonincreasing .}$$

Note: c1 and c2 are obvious. c3 follows from (3.1.4) and c4 is a consequence of (3.1.5).

On the other hand, if h* satisfies c1–c4, let G* be defined by

$$1 - G^*(\theta) = \frac{\int_\theta^\infty \frac{h^*(y)}{\sqrt{y^2-\theta^2}}\,dy}{\int_0^\infty \frac{1}{y}\,h^*(y)dy} \quad .$$

c2–c4 imply that g* is a distribution function with $G^*(0-) = 0$ and

$$\int_0^\infty \theta\, dG^*(\theta) = \int_0^\infty 1 - G^*(\theta)d\theta$$

$$= \frac{\int_0^\infty \int_\theta^\infty \frac{h^*(y)}{\sqrt{y^2-\theta^2}}\,dy\, d\theta}{\int_0^\infty \frac{1}{y}\,h^*(y)dy}$$

$$= \frac{\int_0^\infty h^*(y) \int_0^y \frac{1}{\sqrt{y^2-\theta^2}}\,d\theta\, dy}{\int_0^\infty \frac{1}{y}\,h^*(y)dy}$$

$$= \frac{\pi}{2}\,\frac{1}{\int_0^\infty \frac{1}{y}\,h^*(y)dy} < \infty \; .$$

By reversing the steps in proving (3.1.5) (see Coleman, 1979, P.53), we can prove

$$h^*(y) = \int_y^\infty \frac{y}{\theta\sqrt{\theta^2-y^2}}\, dG^*(\theta|\uparrow)$$

where

$$dG^*(\theta|\uparrow) = \frac{\theta\, dG^*(\theta)}{\int_0^\infty \theta\, dG^*(\theta)} .$$

Therefore c1–c4 is a characterization of $\{h(\cdot|\uparrow)\}$.

Anderssen and Jakeman explained the poor performance of back substitution by the inability of the sample c.d.f. H_n to cope with the singularity of the integrand $1/\sqrt{y^2-\theta^2}$ at $y = \theta$. However, a more basic problem is that $H_n \notin \{H(\cdot|\uparrow)\}$, to see this, note that H_n is discrete while $H(\cdot|\uparrow)$ has density $h(\cdot|\uparrow)$. In this light, even though Anderssen and Jakeman's product integration estimate $\tilde{H}_n$ copes with the singularity, there is no reason to believe that it will solve the more basic problem that $\tilde{H}_n$ may not belong to the admissible range $\{H(\cdot|\uparrow)\}$. In the appendix, we consider the simplest kind of product integration, namely, piecewise linear interpolation. We show that there exists G such that $P(H_n \text{ violates c4}) > P > 0$ for n sufficiently large and some $P > 0$. (See A.1 of Appendix).

Note: Saying that $\tilde{H}_n$ violated c4 is the same as saying that $\tilde{G}_n$ is not nondecreasing, that is, $\tilde{G}_n$ is not a distribution function.

3.4. Nonparametric MLE.

3.4.a. Derivation.

The following argument shows that the MLE is not well defined as it stands.

Recall the identification of section 1.6,

$$Q(\theta) = G(\theta|\uparrow)$$

$$f_\theta(y) = h(y|\theta,\uparrow) = \frac{y}{\theta\sqrt{\theta^2-y^2}} \qquad 0 < y < \theta$$

$$f_Q(y) = h(y|\uparrow) = \int f_\theta(y)\, dQ(y) .$$

We want to find $\hat{Q}$ that maximizes $\prod_{i=1}^{n} f_Q(y_i)$.

Consider the equivalent problem of maximizing $\prod_{i=1}^{n} f(y_i)$ subject to $f \in \{f_Q\}$. Let

$$\Gamma = \{(f_\theta(y_1), f_\theta(y_2), \ldots, f_\theta(y_n)),\quad \theta > 0\}$$

$$\Gamma \subset R^n .$$

The problem is then equivalent to maximizing $\prod_{i=1}^{n} f_i$ subject to $(f_1,\ldots,f_n) \in \text{conv}(\Gamma)$ where $\text{conv}(\Gamma)$ = convex hull of Γ. Because of the singularity of $f_\theta(y)$ at $\theta = y$, Γ is not bounded, so $\text{conv}(\Gamma)$ is not bounded and we cannot maximize $\prod_{i=1}^{n} f_i$.

Using the following limiting process, we can get around the difficulty of the singularity. Let $\underset{\sim}{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, $\varepsilon_i > 0$. Let $\hat{Q}(\underset{\sim}{\varepsilon})$ maximize $\prod_{i=1}^{n} f_Q(y_i)$ but only among those Q's which do not put mass on the intervals $[y_i, y_i + \varepsilon_i)$, $i = 1,\ldots,n$. By doing so, we avoid the problem of singularity and Γ is now bounded. Using the fact that as a function of θ, $f_\theta(y) = 0$ for $\theta \le y$ and $f_\theta(y)$ is decreasing for $\theta > y$ it can be shown that $\hat{G}(\underset{\sim}{\varepsilon})$ must necessarily put all its mass at

$\{y_1+\epsilon_1, y_2+\epsilon_2, \ldots, y_n+\epsilon_n\}$ with corresponding masses $P_1(\underset{\sim}{\epsilon}), P_2(\underset{\sim}{\epsilon}), \ldots, P_n(\underset{\sim}{\epsilon})$. Now if $(P_1(\underset{\sim}{\epsilon}), \ldots, P_n(\underset{\sim}{\epsilon})) \to (P_1, \ldots, P_n)$ as $\underset{\sim}{\epsilon} \to 0$ it would seem reasonable to define the MLE $\hat{Q}$ by the discrete distribution which puts mass P_i at y_i, $i = 1, \ldots, n$.

Claim: $P_i(\underset{\sim}{\epsilon}) \to \frac{1}{n}$ as $\underset{\sim}{\epsilon} \to 0$ for all i.

Proof: Let $f_{ij}(\underset{\sim}{\epsilon}) = f_{y_i+\epsilon_i}(y_j)$, we want to find $P_1, P_2, \ldots, P_n$ which maximize $\Pi_{j=1}^{n} (\Sigma_{i=1}^{n} P_i f_{ij})$ subject to $P_i \geq 0$, $P_1 + \cdots + P_n = 1$.

Notice that $f_{ij}(\underset{\sim}{\epsilon}) = 0$ for $j > i$. Using the method of Lagrange multipliers, it reduces to maximizing

$$\begin{aligned} &\log(P_1 f_{11} + P_2 f_{21} + \cdots + P_n f_{n1}) \\ &+ \log(P_2 f_{22} + P_3 f_{32} + \cdots + P_n f_{n2}) \\ &\quad \vdots \\ &+ \log(P_n f_{nn}) \\ &- \alpha (P_1 + \cdots + P_n - 1) . \end{aligned}$$

Taking partial derivative and setting them to zero give

$$\frac{f_{11}}{P_1 f_{11} + \cdots + P_n f_{n1}} - \alpha = 0$$

$$\frac{f_{21}}{P_1 f_{11} + \cdots + P_n f_{n1}} + \frac{f_{22}}{P_2 f_{22} + \cdots + P_n f_{n2}} - \alpha = 0$$

$$\vdots$$

$$\frac{f_{n1}}{P_1 f_{11} + \cdots + P_n f_{n1}} + \frac{f_{n2}}{P_2 f_{22} + \cdots + P_n f_{n2}} + \cdots + \frac{f_{nn}}{P_n f_{nn}} - \alpha = 0$$

$$P_1 + \cdots + P_n = 1 .$$

In the above equations, $\underset{\sim}{\varepsilon}$ is suppressed. Since

$$f_\theta(y) = \begin{cases} \dfrac{y}{\theta\sqrt{\theta^2-y^2}} & 0 < y < \theta \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_{ij} = f_{y_i+\varepsilon_i}(y_j) \quad ,$$

we see that as

$$\underset{\sim}{\varepsilon} \to 0 \qquad f_{ii} \to \infty$$

$$\{f_{ij},\ i > j\} \text{ is uniformly bounded.}$$

Thus as $\underset{\sim}{\varepsilon} \to 0$, the equations become

$$\frac{1}{P_1} - \alpha = 0$$

$$\frac{1}{P_2} - \alpha = 0$$

$$\vdots$$

$$\frac{1}{P_n} - \alpha = 0$$

$$P_1 + \cdots + P_n = 1$$

from which $(P_1(\underset{\sim}{\varepsilon}), P_2(\underset{\sim}{\varepsilon}),\ldots,P_n(\underset{\sim}{\varepsilon})) \to (\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n})$ easily follows.

Therefore the MLE $\hat{Q}$ puts mass $\frac{1}{n}$ at y_i, $i = 1,\ldots,n$.

3.4b. Inconsistency of MLE.

It suffices to show that $F_{\hat{Q}}$ is an inconsistent estimate of F_Q. Let

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{y_i \le y\}} \quad \text{be the sample c.d.f.}$$

By the Glivenko-Cantelli theorem

$$\sup_y |F_n(y) - F_Q(y)| \to 0 \quad \text{with probability } 1.$$

Suppose $F_{\hat{Q}}$ is consistent in the sense that $F_{\hat{Q}} \to F_Q$ with probability 1. Since F_Q is continuous

$$\sup_y |F_{\hat{Q}}(y) - F_Q(y)| \to 0 \quad \text{with probability 1}$$

and so

$$\sup |F_{\hat{Q}}(y) - F_n(y)| \to 0 \quad \text{with probablity } 1.$$

But

$$F_{\hat{Q}}(y) = \sum_{i=1}^{n} \frac{1}{n} F_{y_i}(y)$$

Since $F_\theta(y) = 1$ if $\theta \le y$

$$F_{\hat{Q}}(y) = \sum_{i=1}^{n} \frac{1}{n} I_{\{y_i \le y\}} + \frac{1}{n} \sum_{i=1}^{n} F_{y_i}(y)\, I_{\{y_i > y\}}$$

$$= F_n(y) + \frac{1}{n} \sum_{i=1}^{n} F_{y_i}(y)\, I_{\{y_i > y\}}.$$

Therefore

$$F_{\hat{Q}}(y) - F_n(y) = \frac{1}{n} \sum_{i=1}^{n} F_{y_i}(y)\ I_{\{y_i > y\}} .$$

By the strong law of large numbers

$$F_{\hat{Q}}(y) - F_n(y) \to E(F_Y(y)\ I_{\{Y > y\}}) > 0 \quad \text{where} \quad Y \sim F_Q .$$

This is a contradiction, so $\hat{Q}$ is not a consistent estimate of Q.

Note: It is remarkable that the restricted MLE $F_{\hat{Q}}$ is inconsistent whereas the unrestricted MLE, the sample c.d.f. F_n is consistent. Other instances of this phenomenon can be found in the literature. For example, the restricted MLE for starshaped and star-ordered classes of distribution is inconsistent. (See Barlow, Bartholomew, Bremner and Brunk, P. 257-258).

3.5. Minimum distance methods.

In this section, we will make frequent use of the identification

$$f_\theta(y) \longleftrightarrow h(y|\theta,\uparrow)$$

$$Q(\theta) \longleftrightarrow G(\theta|\uparrow)$$

$$\int F_\theta(y)dQ(\theta) = F_Q(y) \longleftrightarrow H(y|\uparrow) = \int H(y|\theta,\uparrow)dG(\theta|\uparrow)$$

$$f_Q(y) \longleftrightarrow h(y|\uparrow) .$$

3.5a. Description.

i) Deely and Kruse's approach.

$$Y_1, Y_2, \ldots, Y_n \quad \text{i.i.d.} \sim F_{Q_0}$$

where $Q_0 = G_0(\cdot|\dagger)$ and G_0 is the true particle size distribution.

After observing the data $y_1, y_2, \ldots, y_n$ let $\mathcal{G}_n$ be the set of all discrete distributions with weights at $y_1, y_2, \ldots, y_n$. $\hat{Q}_n$ minimizes $\|F_Q - F_n\|$ among all $Q \in \mathcal{G}_n$, where F_n is the sample c.d.f. This is a linear programming problem.

ii) Choi and Bulgren's approach.

$$\hat{Q}_n \quad \text{minimizes} \int (F_Q(y) - F_n(y))^2 \, dF_n(y)$$

among all $Q \in \mathcal{G}_n$. This is a quadratic programming problem.

Note: In Deely and Kruse (1968) $\{\theta_{n,1}, \theta_{n,2}, \ldots, \theta_{n,n}\}$ are prespecified. In Choi and Bulgren (1968), $\{\theta_{n,1}, \ldots, \theta_{n,n}\}$ are part of the argument of the function we seek to minimize. But here $\{\theta_{n,1}, \ldots, \theta_{n,n}\} = \{y_1, \ldots, y_n\}$ equals the observed data.

3.5b. Consistency of $\hat{G}_n(\theta|\dagger)$.

We need the following lemma.

Lemma 3.5.1. Let

$$Y_1, Y_2, \ldots, Y_n, \ldots \quad \text{i.i.d.} \sim h_0(y|\dagger) = \int_y^\infty \frac{y}{\theta\sqrt{\theta^2 - y^2}} \, dG_0(\theta|\dagger) .$$

For almost every sample sequence $\{y_k\}$ there exists a $Q_n^* \in \mathcal{G}_n$ for each n such that

$$Q_n^* \xrightarrow{\mathcal{D}} Q_0 \quad \text{and} \quad F_{Q_n^*} \xrightarrow{\mathcal{D}} F_{Q_0}$$

Proof. See A.2 of the Appendix

i) Consistency of Deely and Kruse's estimator. With the aid of Lemma 3.5.1, the same proof in their paper goes through.

ii) Consistency of Choi and Bulgren's estimator. By Lemma 3.5.1, for almost every sample sequence $\{y_k\}$, there exists $Q_n^* \in \mathcal{G}_n$ such that $\|F_{Q_n^*} - F_{Q_0}\| \to 0$ (since F_{Q_0} is continuous)

$$\Rightarrow \|F_{Q_n^*} - F_n\| \to 0 \qquad \text{where } F_n \text{ is the sample c.d.f.}$$

$$\Rightarrow \int (F_{Q_n^*}(y) - F_n(y))^2 \, dF_n(y)$$

$$\leq \|F_{Q_n^*} - F_n\|^2 \to 0 \, .$$

By definition

$$\int (F_{\hat{Q}_n}(y) - F_n(y))^2 \, dF_n(y) < \int (F_{Q_n^*}(y) - F_n(y))^2 dF_n(y) \to 0 \, .$$

By the lemma on P. 453, Choi and Bulgren (1968)

$$(\|F_{\hat{Q}_n} - F_n\| - \tfrac{1}{n})^3 \leq 3 \int (F_{\hat{Q}_n}(y) - F_n(y))^2 dF_n(y) \to 0$$

$$\Rightarrow \|F_{\hat{Q}_n} - F_n\| \to 0$$

$$\Rightarrow \|F_{\hat{Q}_n} - F_{Q_0}\| \to 0 \, .$$

By Theorem 2, Robbins (1964)

$$\hat{Q}_n \xrightarrow{\mathcal{L}} Q_0 \, .$$

3.5c. A difficulty in proving consistency of $\hat{G}_n$.

We proved that $\hat{Q}_n(\cdot) = \hat{G}_n(\cdot|\uparrow)$ is consistent in the sense that with probability 1, $\hat{G}_n(\cdot|\uparrow) \xrightarrow{\mathcal{L}} G_0(\cdot|\uparrow)$. By (3.1.2)

$$1 - G_0(\theta-) = \frac{\int_\theta^\infty \frac{1}{\xi}\, dG_0(\xi|\uparrow)}{\int_0^\infty \frac{1}{\xi}\, dG_0(\xi|\uparrow)}$$

So $$1 - \hat{G}_n(\theta-) = \frac{\int_\theta^\infty \frac{1}{\xi}\, d\hat{G}_n(\xi|\uparrow)}{\int_0^\infty \frac{1}{\xi}\, d\hat{G}_n(\xi|\uparrow)} .$$

Since $\frac{1}{\xi}$ is continuous and bounded on $[\theta,\infty)$

$$\hat{G}_n(\cdot|\uparrow) \xrightarrow{\mathcal{L}} G_0(\cdot|\uparrow) \Rightarrow \int_\theta^\infty \frac{1}{\xi}\, d\hat{G}_n(\xi|\uparrow)$$

$$\to \int_\theta^\infty \frac{1}{\xi}\, dG_0(\xi|\uparrow) .$$

Thus the numerator of $1-\hat{G}_n$ converges to the numerator of $1-G_0$. But $\frac{1}{\xi}$ has a singularity at $\xi = 0$ so that $\hat{G}_n(\cdot|\uparrow) \xrightarrow{\mathcal{L}} G_0(\cdot|\uparrow)$ does not necessarily imply

$$\int_0^\infty \frac{1}{\xi}\, d\hat{G}_n(\xi|\uparrow) \to \int_0^\infty \frac{1}{\xi}\, dG_0(\xi|\uparrow) .$$

Thus the denominator may not converge to the right value. Therefore $\hat{G}_n(\cdot|\uparrow) \xrightarrow{\mathcal{L}} G_0(\cdot|\uparrow)$ does not necessarily imply $\hat{G}_n \xrightarrow{\mathcal{L}} G_0$.

3.5d. Modification leading to consistent estimate of G_0.

In the previous section, we see that the problem is that we cannot insure

$$(3.5.1) \qquad \int_0^\infty \frac{1}{\xi}\, d\hat{G}_n(\xi|\uparrow) \to \int_0^\infty \frac{1}{\xi}\, dG_0(\xi|\uparrow) \quad .$$

If (3.5.1) can be insured, then $\hat{G}_n(\cdot|\uparrow) \xrightarrow{\mathcal{L}} G_0(\cdot|\uparrow) \Rightarrow \hat{G}_n \xrightarrow{\mathcal{L}} G_0$. This motivates the following procedures

i) Modified Deely and Kruse

$$\hat{Q}_n \quad \text{minimize} \quad \|F_Q - F_n\| + \left|\int_0^\infty \frac{1}{\xi}\, dQ(\xi) - \frac{2}{\pi}\,\frac{1}{n}\sum_{i=1}^n \frac{1}{y_i}\right|$$

among all $Q \in \mathcal{Q}_n$. Again, this is a linear programming problem.

ii) Modified Choi and Bulgren.

$$\hat{Q}_n \text{ minimizes } \int (F_Q(y)-F_n(y))^2 dF_n(y) + \left(\int_0^\infty \frac{1}{\xi}\, dQ(\xi) - \frac{2}{\pi}\frac{1}{n}\sum_{i=1}^n \frac{1}{y_i}\right)^2 .$$

This is a quadratic programming problem.

To prove consistency results, we need the following:

Fact 3.5.1.

$$\frac{2}{\pi}\,\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i} \to \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi) \quad \text{with probability } 1 .$$

__Proof.__ By (3.1.2) and (3.1.4)

$$\int_0^\infty \frac{1}{\xi}\, dQ_0(\xi) = \int_0^\infty \frac{1}{\xi}\, dG_0(\xi|\uparrow) = \frac{2}{\pi}\int_0^\infty \frac{1}{y}\, h_0(y|\uparrow)dy ,$$

so that fact 3.5.1 follows from the strong law of large numbers.

Lemma 3.5.2. For almost every sample sequence, there exists a $Q_n^* \in \mathcal{G}_n$ for each n such that

$$Q_n^* \xrightarrow{\mathcal{B}} Q_0 \ , \quad F_{Q_n^*} \xrightarrow{\mathcal{B}} F_{Q_0}$$

and

$$\int_0^\infty \frac{1}{\xi}\, dQ_n^*(\xi) \to \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi) \quad .$$

Proof. See A.2 of Appendix.

We can now prove the consistency of $\hat{G}_n$. For the modified Deely and Kruse's procedure, by definition

$$\|F_{Q_n} - F_n\| + \left|\int_0^\infty \frac{1}{\xi}\, d\hat{Q}_n(\xi) - \frac{2}{\pi}\frac{1}{n}\sum_{i=1}^n \frac{1}{y_i}\right|$$

$$< \|F_{Q_n^*} - F_n\| + \left|\int_0^\infty \frac{1}{\xi}\, dQ_n^*(\xi) - \frac{2}{\pi}\frac{1}{n}\sum_{i=1}^n \frac{1}{y_i}\right|$$

$\to 0$ for almost sample sequence $\{y_k\}$ by fact 3.5.1 and lemma 3.5.2.

So $\|F_{\hat{Q}_n} - F_n\| \to 0$ which implies $\|F_{\hat{Q}_n} - F_{Q_0}\| \to 0$. By Theorem 2, Robbins (1964) $\hat{Q}_n \xrightarrow{\mathcal{B}} Q_0$. Also we have

$$\left|\int_0^\infty \frac{1}{\xi}\, d\hat{Q}_n(\xi) - \frac{2}{\pi}\frac{1}{n}\sum_{i=1}^n \frac{1}{y_i}\right| \to 0$$

So

$$\int_0^\infty \frac{1}{\xi}\, d\hat{Q}_n(\xi) \to \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi)$$

by fact 3.5.1. In other words, (3.5.1) is satisfied. Together, these imply $\hat{G}_n \xrightarrow{\mathcal{E}} G_0$ with probability 1.

The proof of the consistency of the modified Choi and Bulgren's procedure is analogous and will be omitted.

3.5e. Comparison with existing methods.

i) In the next section we will present simulation results which are indicative of the better performance of the minimum distance method.

ii) The minimum distance method actually gives an estimate $\hat{G}_n$ of G_0. The back substitution estimate G_n is not even a d.f. There is no reason to believe that the product integration estimate $\tilde{G}_n$ is a d.f. either.

iii) An equivalent way to express ii) is that the minimum distance method takes into account the structure of the problem so that $\hat{F}_n = F_{\hat{Q}_n} \in \{F_Q\}$ whereas F_n and $\tilde{F}_n \notin \{F_Q\}$. The necessity to have an estimate belong to $\{F_Q\}$ is magnified when we come to the section on the bootstrap (section 3.7).

3.6. Simulation results.

Case 1. $Q_0(\theta) = G_0(\theta|\uparrow)$ puts weight 0.2 at $\theta = 1,2,3,4,5$. Sample size $n = 100$.

Note: i) We assume that the support points 1,2,3,4,5 are known so that we have a four parameter problem instead of an infinite parameter problem as in the fully nonparametric case.

ii) In this case of a finite mixture with known support, the MLE will be consistent.

iii) The minimum distance method is implemented without the modification described in section 3.5d.

iv) In computing Choi and Bulgren's estimate, we do not need to solve a quadratic programming problem. Without the nonnegativity constraint the solution is just the solution of a linear system of $5-1 = 4$ equations. The solution to the linear system will usually satisfy the nonnegativity constraint automatically because we expect the estimate to be near the true value of 0.2 as the sample size increases.

v) Let

$$\mu = \int_0^\infty \theta \, dG(\theta) . \tag{3.6.1}$$

By (3.1.4)

$$\mu = \frac{\pi}{2} \frac{1}{\int_0^\infty \frac{1}{y} h(y|\uparrow) \, dy} . \tag{3.6.2}$$

MLE and minimum distance method estimate μ by replacing $dG(\theta)$ by $d\hat{G}_n(\theta)$ in (3.6.1). Back substitution and product integration estimate μ by replacing $h(y|\uparrow)\,dy$ in (3.6.2) by $dH_n(y)$ and $d\tilde{H}_n(y)$ respectively.

The results are summarized in tables 3,4,5.

<u>Case 2</u>. $G_0(\theta)$ puts mass 0.2 at $\theta = 1,2,3,4,5$. Sample size $n = 100$. The results are summarized in tables 6,7,8.

Note: In cases 1 and 2, we have cheated a little bit for we assume we know the support points $\theta = 1,2,3,4,5$. When they are not known, the nonparametric MLE is inconsistent (section 3.4). However, the minimum distance estimate is still consistent (section 3.5).

In spite of this, we can still say the following:

i) The simulation results are still indicative of better performance by the minimum distance method over existing methods since it takes into account the structure of the problem. This is confirmed in case 3.

ii) For the thin slice case, if G_0 is discrete with finite support, we can in fact know the points of support from the data since at those values, we should have repeated measurements with high probability for large samples. (see sections 4.5,4.8).

<u>Case 3</u>. $$g(\theta) = \frac{\theta}{c^2} \exp(-\frac{\theta^2}{2c^2}) , \qquad \theta > 0 ,$$

that is, the stationary distribution, $c = 4$, $n = 100$.

Note: i) This is a "true" test of the minimum distance method since we compute it "honestly" here without the incorporation of additional information concerning G.

ii) We use Deely and Kruse's approach.

iii) After averaging over 100 trials, Anderssen and Jakeman's estimate $\tilde{G}_n$ seems to be increasing. But in a substantial number of the 100 trials, $\tilde{G}_n$ is not increasing. They are not reproduced here for lack of space. The results summarized in tables 9 and 10 are still in favor of the minimum distance method.

$G(\theta|\uparrow)$ puts mass 0.2 at $\theta = 1,2,3,4,5$

Sample size = 100

planar data

TABLE 3

The average over 100 trials of the Choi and Bulgren estimate and the maximum likelihood estimate of the probability mass functions $P(\theta)$ and $P(\theta|\uparrow)$ and their m.s.e. (shown on even rows).

θ	$P(\theta)$	C&B	MLE	$P(\theta\|\uparrow)$	C&B	MLE
1.0000	0.4380	0.4283	0.4186	0.2000	0.1977	0.1906
		0.0075	0.0059		0.0027	0.0021
2.0000	0.2190	0.2215	0.2264	0.2000	0.2008	0.2032
		0.0058	0.0043		0.0046	0.0034
3.0000	0.1460	0.1499	0.1505	0.2000	0.2026	0.2020
		0.0029	0.0023		0.0048	0.0037
4.0000	0.1095	0.1152	0.1147	0.2000	0.2083	0.2049
		0.0017	0.0011		0.0051	0.0031
5.0000	0.0876	0.0851	0.0898	0.2000	0.1907	0.1993
		0.0010	0.0007		0.0041	0.0027

TABLE 4

A comparison between Anderssen and Jakeman's estimate, Choi and Bulgren's estimate and the maximum likelihood estimate of G(θ) in terms of m.s.e. (shown on even rows) and the average of the estimates over 100 trials.

θ	G(θ)	A&J	C&B	MLE
0.5000	0.0000	-0.0658	0.0000	0.0000
1.0000	0.4380	0.3100	0.4283	0.4186
		32.1618	15.5362	14.1427
1.5000	0.4380	0.3738	0.4283	0.4186
		24.3492	15.5362	14.1427
2.0000	0.6569	0.5802	0.6498	0.6450
		14.7763	7.9217	6.8735
2.5000	0.6569	0.6211	0.6498	0.6450
		10.5440	7.9217	6.8735
3.0000	0.8029	0.7382	0.7997	0.7956
		9.7252	4.4107	4.2060
3.5000	0.8029	0.7872	0.7997	0.7956
		5.5502	4.4107	4.2060
4.0000	0.9124	0.8877	0.9149	0.9102
		4.0513	2.7474	2.3778
4.5000	0.9124	0.9104	0.9149	0.9102
		3.1911	2.7474	2.3778
5.0000	1.0000	1.0000	1.0000	1.0000
		0.0000	0.0000	0.0000

TABLE 5

A comparison of the back substitution estimate, Anderssen and Jakeman's estimate, Choi and Bulgren's estimate and the maximum likelihood estimate of the mean sphere radius μ in terms of m.s.e. (shown on second row) and the average of the estimates over 100 trials.

μ	B.S.	A & J	C & B	MLE
2.1898	2.2850	2.3965	2.2073	2.2305
	0.0771	0.1067	0.0302	0.0271

Note: B.S. = Back substitution, A & J = Anderssen & Jakeman, C&B = Choi and Bulgren

G(θ) puts mass 0.2 at θ = 1,2.3.4.5

Sample size = 100

Planar data

TABLE 6

The average over 100 trials of the Choi and Bulgren estimate and the maximum likelihood estimate of the probability mass functions $P(\theta)$ and $P(\theta|\uparrow)$ and their m.s.e. (shown on even rows).

θ	$P(\theta)$	C&B	MLE	$P(\theta\|\uparrow)$	C&B	MLE
1.0000	0.2000	0.2262	0.2161	0.0667	0.0810	0.0751
		0.0145	0.0080		0.0024	0.0013
2.0000	0.2000	0.1930	0.1927	0.1333	0.1316	0.1303
		0.0096	0.0052		0.0042	0.0023
3.0000	0.2000	0.2017	0.1992	0.2000	0.2063	0.2013
		0.0073	0.0060		0.0072	0.0058
4.0000	0.2000	0.1865	0.1916	0.2667	0.2540	0.2576
		0.0039	0.0031		0.0059	0.0048
5.0000	0.2000	0.1926	0.2003	0.3333	0.3271	0.3357
		0.0023	0.0022		0.0043	0.0041

TABLE 7

A comparison between Anderssen and Jakeman's estimate, Choi and Bulgren's estimate and the maximum likelihood estimate of $G(\theta)$ in terms of m.s.e. (shown on even rows) and the average of the estimates over 100 trials.

θ	$G(\theta)$	A&J	C&B	MLE
0.5000	0.0000	-0.0613	0.0000	0.0000
1.0000	0.3000	0.1269	0.2262	0.2161
		59.1435	47.2917	34.6940
1.5000	0.2000	0.1655	0.2262	0.2161
		64.7058	47.2917	34.6940
2.0000	0.4000	0.3004	0.4192	0.4089
		37.2330	18.1612	15.1627
2.5000	0.4000	0.3764	0.4192	0.4089
		25.3233	18.1612	15.1627
3.0000	0.6000	0.5397	0.6210	0.6081
		16.9577	9.5857	8.8996
3.5000	0.6000	0.5915	0.6210	0.6081
		11.6811	9.5857	8.8996
4.0000	0.8000	0.7670	0.8074	0.7997
		7.6166	4.7007	4.5665
4.5000	0.8000	0.7859	0.8074	0.7997
		7.0196	4.7007	4.5665
5.0000	1.0000	1.0000	1.0000	1.0000
		0.0000	0.0000	0.0000

TABLE 8

A comparison of the back substitution estimate, Anderssen and Jakeman's estimate, Choi and Bulgren's estimate and the maximum likelihood estimate of the mean sphere radius μ in terms of m.s.e. (shown on second row) and the average of the estimates over 100 trials.

μ	B.S.	A&J	C&B	MLE
3.0000	2.9403	3.1430	2.9262	2.9673
	0.2022	0.1369	0.0751	0.0489

$g(\theta) = \frac{\theta}{16} e^{-\theta^2/32}$, the stationary distribution

Sample size = 100

Planar data

TABLE 9

A comparison between Anderssen and Jakeman's estimate and Deely and Kruse's estimate of $G(\theta)$ in terms of m.s.e. (shown on even rows) and the average of the estimates over 100 trials.

θ	$G(\theta)$	A&J	D&K
1.2812	0.0500	0.0209	0.0978
		278.295	163.940
1.8362	0.1000	0.0451	0.1230
		138.269	95.6568
2.2805	0.1500	0.1088	0.1635
		71.2289	53.2543
2.6722	0.2000	0.1344	0.2089
		59.7322	45.6907
3.0341	0.2500	0.2296	0.2599
		48.6981	39.2419
3.3784	0.3000	0.2596	0.3139
		35.3034	28.1828
3.7128	0.3500	0.3352	0.3589
		23.2890	22.2477
4.0431	0.4000	0.3663	0.3933
		30.8967	23.5758
4.3739	0.4500	0.3775	0.4267
		34.1915	24.7209
4.7096	0.5000	0.4472	0.4903
		26.0370	22.0270
5.0549	0.5500	0.4884	0.5429
		31.3737	20.9288
5.4149	0.6000	0.5622	0.5871
		16.8953	11.8931
5.7961	0.6500	0.6039	0.6239
		13.9161	11.6400
6.2070	0.7000	0.6828	0.6906
		10.5103	10.5589
6.6604	0.7500	0.6928	0.7345
		15.0637	9.2468
7.1765	0.8000	0.7783	0.7847
		7.7646	5.9378
7.7915	0.8500	0.8226	0.8244
		7.3098	6.3145
8.5839	0.9000	0.8986	0.8964
		2.9213	2.7481
9.7910	0.9500	0.9413	0.9379
		3.1622	2.7778

TABLE 10

A comparison of the back substitution estimate, Anderssen and Jakeman's estimate and Deely and Druse's estimate of the mean sphere radius μ in terms of m.s.e. (shown on second row) and the average of the estimates over 100 trials.

μ	B.S.	A&J	D&K
5.0133	5.0382	5.3666	5.0068
	1.0194	0.8847	0.4658

Note: D&K = Deely and Kruse.

3.7. Bootstrap.

3.7a. Why is it necessary?

The distributions of stereological estimators are notoriously hard to find. Take the simplest case of

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{Y_i} ;$$

since $E(\frac{1}{Y^2}) = \infty$, T_n has infinite variance. Watson (1971) showed that T_n is asymptotically normal, but the approximation is not usable for two reasons: i) the normalizing factor and asymptotic variance depends on whether

$$\int_0^\infty \frac{1}{\theta}\, dG(\theta) < \infty$$

or not.

ii) Monte Carlo experiments showed that the normal approximation was very poor when $n = 100$.

Bloomfield showed that by simply omitting the first term of the classical estimator, the resulting estimator $\frac{1}{n}\Sigma_{i=2}^{n}\frac{1}{Y_{n,i}}$ has finite variance even though it is biased.

Jakeman and Schaeffer (1978) found conditions under which

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i} \quad \text{and} \quad \frac{1}{n}\sum_{i=2}^{n}\frac{1}{Y_{n,i}}$$

have the same asymptotic distribution. Since Anderssen and Jakeman's trapezoidal product integration estimate is sandwiched in between the above two estimates, it also has the same asymptotic distribution.

Jakeman and Schaeffer also simulated 50 independent samples of size 150, 300, and 2000. They compare the theoretical asymptotic variance with the variance obtained from the fifty samples. Three distributions are tried. For the classical estimate

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i},$$

there is no agreement even with $n = 2000$. For Anderssen and Jakeman's estimate, there is reasonable agreement for two of the three distributions but only at $n = 2000$.

In conclusion, stereologists so far have not been able to find a usable and satisfactory approximation to the distribution of a stereological estimate. As a result, they cannot do any inference.

3.7b. <u>Introduction</u>.

In an intuitive and non rigorous manner, we will indicate how the bootstrap can be profitably used. Also, we will indicate the wrong way and the right way to bootstrap.

Recall that

$$Y_1, Y_2, \ldots, Y_n \quad \text{i.i.d.} \sim F_Q = \int F_\theta \, dQ(\theta) \ .$$

Since $E(\frac{1}{Y^2}) = \infty$, the variance is infinite so that the usual central limit theorem does not hold. Watson (1971) shows that

i) If $\int_0^\infty \frac{1}{\theta} \, dG(\theta) < \infty$

$$\sqrt{\frac{n}{\log n}} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{Y_i} - \beta\right) \to N\left(0, \frac{1}{2\mu} \int_0^\infty \frac{1}{\theta} \, dG(\theta)\right)$$

where Q and G are related by $Q(\theta) = G(\theta|\uparrow)$, (Section 1.6) and

$$\beta = E\left(\frac{1}{Y}\right) = \frac{\pi}{2\mu}$$

and

$$\mu = \int_0^\infty \theta \, dG(\theta) \ .$$

ii) If $\int_0^\infty \frac{1}{\theta} \, dG(\theta) = \infty$, then

$$\frac{\sqrt{n}}{\log n} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{Y_i} - \beta\right) \to N\left(0, \frac{g(0)}{8\mu}\right) \ .$$

Usually, when we bootstrap, we resample $Y_1^*, Y_2^*, \ldots, Y_n^*$ from F_n, the sample c.d.f. In fact, in many cases, we can resample from any consistent estimate. However, in our case, as we shall see shortly, we have to be more careful regarding which estimate of F to use.

3.7c. The wrong way to bootstrap.

Resampling from the sample c.d.f. F_n does not work. Intuitively, the reason is the following: If $Y_1^*, Y_2^*, \ldots, Y_n^*$ i.i.d. $\sim F_n$ then $E(\frac{1}{Y^{*2}}) < \infty$. So the usual central limit theorem holds. The normalizing constant would be $\sqrt{n}$ instead of $\sqrt{\frac{n}{\log n}}$ or $\frac{\sqrt{n}}{\log n}$. Thus we would not expect the bootstrap to work. This is indeed supported by simulation results which we will discuss later.

3.7d. The right way to bootstrap.

Assume $\int_0^\infty \frac{1}{\theta} dG(\theta) < \infty$, we resample from a consistent estimate $\hat{F}_n \in \{F_Q\}$ (this is the case for the minimum distance method).

Let $\hat{F}_n = F_{\hat{Q}_n}$.

Since $\hat{Q}_n$ is typically discrete with finite support, $\int_0^\infty \frac{1}{\theta} d\hat{G}_n(\theta) < \infty$. Thus if $Y_1^*, Y_2^*, \ldots, Y_m^*$ i.i.d. $\sim F_{\hat{Q}_n}$, then for each fixed n, as $m \to \infty$

$$\sqrt{\frac{m}{\log m}} \left(\frac{1}{m} \sum_{i=1}^{m} \frac{1}{Y_i^*} - E(\frac{1}{Y^*})\right) \to N(0, \frac{1}{2\mu^*} \int_0^\infty \frac{1}{\theta} d\hat{G}_n(\theta))$$

where

$$\mu^* = \int_0^\infty \theta \, d\hat{G}_n(\theta).$$

Hence it is plausible that bootstrapping in this manner will work. Simulation results seem to support this. The above argument is not rigorous; the missing step is some kind of uniformity argument which enables us to go from (fixed n, $m \to \infty$) to ($m(n) = n$, $n \to \infty$). This step is often the essential but hardest step in proving results concerning the bootstrap.

But what makes the bootstrap attractive despite the limited theoretical results is the large number of examples where the bootstrap seems to work, including complicated situations where there is no standard solution. Besides suggesting another application of the bootstrap, our present example illustrates another point, namely, we have to be careful in bootstrapping, contrary to what it appears to be, "Bootstrap" is not a cookbook method.

Note: We assume $\int_0^\infty \frac{1}{\theta}\,dG(\theta) < \infty$ which admittedly is a limitation. However when there is truncation (only $y \geq y_0$ is observed) this condition is automatically satisfied. For details, see section 3.8.

3.7e. Simulation results.

Case 1. $G(\theta|\uparrow)$ puts mass 0.2 at $\theta = 1,2,3,4,5$. Sample size $n = 100$.

Let

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{Y_i} - E(\frac{1}{Y})$$

$$T_2 = \frac{1}{n-1}\sum_{i=2}^{n} \frac{1}{Y_{n,i}} - E(\frac{1}{Y}) .$$

For both T_1 and T_2, we bootstrap the following quantities: variance and the five quantiles $Q_{.10}$, $Q_{.25}$, $Q_{.50}$, $Q_{.75}$ and $Q_{.90}$.

i) The "true" values which are obtained by generating 1,000 independent samples of size 100 are as follows:

For T_1

var = 0.0133

$Q_{.10} = -0.1146$ $\quad Q_{.25} = -0.0730$ $\quad Q_{.50} = -0.0260$

$Q_{.75} = 0.0352$ $\quad Q_{.90} = 0.1067$

For T_2

$$\text{var} = 0.0042$$

$$Q_{.10} = -0.1445 \qquad Q_{.25} = -0.1072 \qquad Q_{.50} = -0.0718$$

$$Q_{.75} = -0.0281 \qquad Q_{.90} = 0.0217$$

ii) Watson's asymptotic theory. If

$$\int_0^\infty \frac{1}{\theta}\, dG(\theta) < \infty$$

then
$$\sqrt{\frac{n}{\log n}}\, T_1 \to N\left(0, \frac{1}{2\mu}\int_0^\infty \frac{1}{\theta}\, dG(\theta)\right)$$

$$\sqrt{\frac{n}{\log n}}\, T_2 \to N\left(0, \frac{1}{2\mu}\int_0^\infty \frac{1}{\theta}\, dG(\theta)\right) .$$

In our case

$$\int_0^\infty \frac{1}{\theta}\, dG(\theta) = 0.6410 \ , \quad \mu = 2.1898 \ .$$

From these, an estimate of var T_1 as well as var T_2 is 0.00674. Estimates of quantiles of both T_1, T_2 are -0.1051, -0.0554, 0, 0.0554, 0.1051. Thus the normal approximation is unsatisfactory, especially T_2.

iii) Resample from the sample c.d.f. (the wrong way to bootstrap). The results are summarized in Tables 11, 12. We can see that the bootstrap estimates vary greatly from trial to trial and they do not seem to center around the true value. Thus simulation is in agreement with the conclusion of section 3.7c; resampling from the sample c.d.f. does not work.

iv) Resample from MLE (the right way).

Note: It may be more fitting to choose minimum distance estimates instead of the MLE. However, the conjecture of section 3.7d is that we can resample from a consistent estimate $\hat{F}_n$ satisfying $\hat{F}_n \varepsilon \{F_Q\}$. And in our present case of finite mixture with known support, the MLE is consistent. So the MLE will serve equally well the purpose of verifying the conjecture numerically. The results are summarized in tables 13 and 14.

The results are not spectacular, but are much better than when we resample from the sample c.d.f. The estimates are much more stable, they do not vary that much from trial to trial. As to whether they center around the true values, the bootstrap is quite successful for T_2, while it seems to fail for $var(T_1)$ but this is to be expected since T_1 has infinite variance even though it has finite asymptotic variance.

v) We attempt to find out whether the failure of resampling from the sample c.d.f. F_n is due to (i) $F_n \not\in \{F_Q\}$, the proposed explanation, or just (ii) $E(\frac{1}{Y^*}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}$ has infinite variance.

We throw away $y_{n,1}$ and resample from the resulting sample c.d.f., then $E(\frac{1}{Y^*}) = \frac{1}{n-1}\sum_{i=2}^{n}\frac{1}{Y_{n,i}}$ has finite variance.

The results are in tables 15 and 16, even though there are only two trials, it is quite evident that the bootstrap does not work well in this case. Thus it appears the problem is due to $F_n \not\in \{F_Q\}$.

TABLE 11

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T_1 = \frac{1}{n}\Sigma_{i=1}^{n}\frac{1}{Y_i} - \beta$, 10 trials, sample size = 100. $G(\theta|\uparrow)$ puts mass 0.2 at θ = 1,2,3,4,5 with resampling from the sample c.d.f.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$
0.0133	-0.1146	-0.0730	-0.0260	0.0352	0.1067

BOOTSTRAP ESTIMATES (resample 1000 times from the sample c.d.f.)

0.0042	-0.0788	-0.0435	-0.0015	0.0422	0.0857
0.0130	-0.1386	-0.0791	-0.0083	0.0740	0.1512
0.0027	-0.0659	-0.0347	-0.0017	0.0339	0.0667
0.0179	-0.1563	-0.0925	-0.0102	0.0868	0.1811
0.0025	-0.0616	-0.0326	-0.0010	0.0335	0.0665
0.0021	-0.0577	-0.0230	-0.0022	0.0310	0.0608
0.0064	-0.0945	-0.0552	-0.0053	0.0506	0.1062
0.0196	-0.1591	-0.1152	-0.0130	0.0849	0.1883
0.0016	-0.0501	-0.0271	-0.0016	0.0258	0.0513
0.0022	-0.0599	-0.0300	0.0005	0.0331	0.0633

TABLE 12

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T_2 = \frac{1}{n-1}\ \Sigma_{i=2}^{n}\frac{1}{Y_{n,i}} - \beta$, 10 trials, sample size 100. $G(\theta|\uparrow)$ puts mass 0.2 at θ = 1,2,3,4,5 with resampling from the sample c.d.f.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$
0.0042	-0.1445	-0.1072	-0.0718	-0.0281	0.0217

BOOTSTRAP ESTIMATES (resample 1000 times from the sample c.d.f.)

0.0037	-0.1068	-0.0744	-0.0366	0.0071	0.0478
0.0117	-0.1930	-0.1389	-0.0756	0.0040	0.0809
0.0024	-0.0875	-0.0587	-0.0276	0.0074	0.0372
0.0148	-0.2226	-0.1670	-0.0978	-0.0087	0.0849
0.0023	-0.0838	-0.0551	-0.0234	0.0106	0.0421
0.0020	-0.0774	-0.0519	-0.0247	0.0086	0.0374
0.0051	-0.1317	-0.0970	-0.0554	-0.0081	0.0444
0.0114	-0.1888	-0.1609	-0.1231	-0.0472	0.0552
0.0014	-0.0665	-0.0465	-0.0223	0.0032	0.0276
0.0022	-0.0797	-0.0492	-0.0191	0.0139	0.0434

TABLE 13

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T_1 = \frac{1}{n} \Sigma_{i=1}^{n} \frac{1}{Y_i} - \beta$, 10 trials, sample size = 100, $G(\theta|\uparrow)$ puts mass 0.2 at θ = 1,2,3,4,5 with resampling from the MLE.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.90}$
0.0133	-0.1146	-0.0730	-0.0260	0.0352	0.1067

BOOTSTRAP ESTIMATES (resample 1000 times from the MLE)

0.0228	-0.1081	-0.0722	-0.0210	0.0523	0.1283
0.0225	-0.1057	-0.0687	-0.0216	0.0511	0.1261
0.0197	-0.0940	-0.0616	-0.0172	0.0464	0.1090
0.0235	-0.1147	-0.0749	-0.0189	0.0552	0.1276
0.0217	-0.1049	-0.0652	-0.0163	0.0409	0.1248
0.0173	-0.0814	-0.0560	-0.0137	0.0406	0.0981
0.0202	-0.1000	-0.0643	-0.0189	0.0481	0.1131
0.0198	-0.0979	-0.0610	-0.0161	0.0447	0.1122
0.0180	-0.0891	-0.0597	-0.0166	0.0420	0.1070
0.0230	-0.1072	-0.0709	-0.0200	0.0525	0.1317

AVERAGE OVER 10 TRIALS

0.0208	-0.1003	-0.0654	-0.0180	0.0472	0.1178

TABLE 14

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T_2 = \frac{1}{n-1} \Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$, 10 trials, sample size = 100, $G(\theta|\uparrow)$ puts mass 0.2 at θ = 1,2,3,4,5 with resampling from the MLE.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$
0.0042	-0.1445	-0.1072	-0.0718	-0.0281	0.0217

BOOTSTRAP ESTIMATES (resample 1000 times from the MLE)

0.0050	-0.1428	-0.1114	-0.0695	-0.0180	0.0273
0.0048	-0.1374	-0.1080	-0.0688	-0.0195	0.0280
0.0037	-0.1265	-0.0967	-0.0615	-0.0182	0.0269
0.0052	-0.1501	-0.1151	-0.0729	-0.0205	0.0284
0.0044	-0.1398	-0.1071	-0.0669	-0.0212	0.0252
0.0028	-0.1110	-0.0858	-0.0555	-0.0159	0.0191
0.0040	-0.1315	-0.1018	-0.0634	-0.0202	0.0239
0.0037	-0.1284	-0.0993	-0.0624	-0.0201	0.0250
0.0031	-0.1162	-0.0930	-0.0588	-0.0179	0.0219
0.0049	-0.1435	-0.1116	-0.0709	-0.0196	0.0255

AVERAGE OVER 10 TRIALS

0.0042	-0.1327	-0.1027	-0.0650	-0.0191	0.0251

TABLE 15

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T_1 = \frac{1}{n}\Sigma_{i=1}^{n} \frac{1}{Y_i} - \beta$, two trials, sample size = 100, $G(\theta|\uparrow)$ puts mass 0.2 at θ = 1,2,3,4,5 with resampling from the sample c.d.f. when $Y_{n,1}$ is thrown away from the sample.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
0.0133	-0.1146	-0.0730	-0.0260	0.0352	0.1067

BOOTSTRAP ESTIMATES (resample 1000 times from the sample c.d.f. when $Y_{n,1}$ is thrown away from the sample).

0.0026	-0.1114	-0.0821	-0.0477	-0.0128	0.0167
0.0079	-0.1893	-0.1406	-0.0859	-0.0238	0.0358

TABLE 16

Bootstrap estimates of the variance and the 0.1, 0.25, 0.75, 0.9 quantiles of $T_2 = \frac{1}{n-1}\Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$, two trials, sample size = 100, $G(\theta|\uparrow)$ puts mass 0.2 at θ = 1,2,3,4,5 with resampling from the sample c.d.f. when $Y_{n,1}$ is thrown away from the sample.

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
0.0042	-0.1445	-0.1072	-0.0718	-0.0281	0.0217

BOOTSTRAP ESTIMATES (resample 1000 times from the sample c.d.f. when $Y_{n,1}$ is thrown away from the sample).

0.0025	-0.1336	-0.1006	-0.0676	-0.0332	-0.0029
0.0067	-0.2298	-0.1875	-0.1368	-0.0810	-0.0233

Case 2. $g(\theta) = \frac{\theta}{c^2} \exp(-\theta^2/2c^2)$, $\theta > 0$. The stationary distribution $c = 4$, $n = 100$.

This case is included since in the preceding case, as in Case 1 of section 3.6, we assume we know the points of support. We bootstrap the same six quantities but only for $T = \frac{1}{n-1} \Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$. Again, the "true" values are obtained by generating 1000 independent samples of size 100.

i) Resample from the sample c.d.f. The results summarized in Table 17 are unsatisfactory.

TABLE 17

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T = \frac{1}{n-1} \Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$, 10 trials, sample size = 100, $g(\theta) = \frac{\theta}{16} e^{-\theta^2/32}$ with resampling from the sample c.d.f.

TRUE VALUES (based on 1000 independent samples).

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
0.000940	-0.062868	-0.049041	-0.033234	-0.014582	0.008415

BOOSTRAP ESTIMATES (resample 1000 times from the sample c.d.f.)

0.003491	-0.102954	-0.086164	-0.057824	-0.012124	0.042565
0.002176	-0.082955	-0.061658	-0.032484	0.000036	0.035518
0.000208	-0.023839	-0.016539	-0.007060	0.003126	0.012128
0.000238	-0.025432	-0.017687	-0.007264	0.003130	0.012771
0.005164	-0.122902	-0.102485	-0.072150	-0.017414	0.045898
0.000537	-0.041454	-0.028469	-0.013718	0.002967	0.017921
0.013839	-0.210793	-0.183493	-0.113879	-0.015015	0.075649
0.000500	-0.040470	-0.028547	-0.013560	0.001731	0.017817
0.001743	-0.078458	-0.061143	-0.035821	-0.006285	0.028975
0.000485	-0.039962	-0.029277	-0.016017	-0.003058	0.013771

ii) Resample from $F_{\hat{Q}_n}$ where $\hat{Q}_n$ is some consistent estimate of Q.

1) $\hat{Q}_n$ = Deely and Kruse's estimate. The results summarized in Table 18 are still unsatisfactory.

We think the failure is due more to insufficient sample size rather than the theory itself. To illustrate this, we consider the following two estimates which should be progressively better.

2) $\hat{Q}_n$ is Deely and Kruse's estimate except that we minimize the distance from F_Q^* instead of F_n where F_Q^* is the true F_Q known only up to its values at $y_1, y_2, \ldots, y_n$.

The results are in Table 19.

3) $\hat{G}_n = F_n$ the sample c.d.f. $\hat{Q}_n(\cdot) = \hat{G}_n(\cdot|\dagger)$.

Since we are sampling from the stationary distribution, $G = F_Q$ so that F_n which estimates F_Q consistently also estimates G consistently.

The results given in Table 20 are quite satisfactory.

TABLE 18

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T = \frac{1}{n-1} \Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$, 10 trials, sample size = 100, $g(\theta) = \frac{\theta}{16} e^{-\theta^2/32}$ with resampling from Deely and Kruse's estimate.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
0.000940	-0.062868	-0.049041	-0.033234	-0.014582	0.008415

BOOTSTRAP ESTIMATES (resample 1000 times from Kruse and Deely's estimate).

0.001708	-0.085659	-0.072052	-0.052141	-0.026198	0.003914
0.001989	-0.090900	-0.073839	-0.050470	-0.018998	0.015578
0.000434	-0.043723	-0.034247	-0.022845	-0.007393	0.007296
0.000565	-0.049936	-0.038930	-0.024875	-0.007648	0.008182
0.001926	-0.089928	-0.075226	-0.055922	-0.029547	0.000282
0.000938	-0.063185	-0.050022	-0.032467	-0.010730	0.013962
0.005940	-0.162466	-0.143522	-0.114927	-0.069700	0.007194
0.000785	-0.056717	-0.045806	-0.028815	-0.010053	0.013531
0.001100	-0.066572	-0.053197	-0.034692	-0.013973	0.012597
0.000504	-0.047352	-0.036867	-0.023304	-0.007649	0.007293

TABLE 19

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T = \frac{1}{n-1} \Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$, 10 trials, sample size = 100, $g(\theta) = \frac{\theta}{16} e^{-\theta^2/32}$ with resampling from Deely and Kruse's estimate when $H = F_Q$ is known up to its values at $y_1, \ldots, y_n$.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
0.000940	-0.062868	-0.049041	-0.033234	-0.014582	0.008415

BOOTSTRAP ESTIMATES (resample 1000 times from Deely and Kruse's estimate when $H = F_Q$ is known up to its values at $y_1, \ldots, y_n$).

0.000872	-0.062096	-0.049394	-0.031613	-0.012375	0.011283
0.000686	-0.054561	-0.042848	-0.026699	-0.008555	0.011001
0.000574	-0.049875	-0.039274	-0.025585	-0.007829	0.009260
0.000581	-0.049992	-0.039118	-0.024924	-0.007506	0.009355
0.000613	-0.051867	-0.040340	-0.025498	-0.007999	0.008047
0.000907	-0.060692	-0.049103	-0.031294	-0.010354	0.013510
0.000641	-0.052962	-0.041608	-0.026311	-0.008401	0.010680
0.000629	-0.052042	-0.041308	-0.025856	-0.008568	0.010410
0.000629	-0.052456	-0.041750	-0.026272	-0.008925	0.009852
0.000801	-0.057060	-0.046070	-0.029845	-0.010247	0.011531

TABLE 20

Bootstrap estimates of the variance and the 0.1, 0.25, 0.5, 0.75, 0.9 quantiles of $T = \frac{1}{n-1} \Sigma_{i=2}^{n} \frac{1}{Y_{n,i}} - \beta$, 10 trials, sample size = 100, $g(\theta) = \frac{\theta}{16} e^{-{}^2/32}$ with resampling from $F_{\hat{Q}}$ where $\hat{Q} = \hat{G}(\cdot|\uparrow)$ and $\hat{G}$ is the sample c.d.f.

TRUE VALUES (based on 1000 independent samples)

Var	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	$Q_{0.9}$
0.000940	-0.062868	-0.049041	-0.033234	-0.014582	0.008415

BOOTSTRAP ESTIMATES (resample 1000 times from $F_{\hat{Q}}$ where $\hat{Q} = \hat{G}(\cdot|\uparrow)$ and $\hat{G}$ is the sample c.d.f.)

0.000835	-0.063257	-0.050242	-0.033450	-0.015003	0.008661
0.000969	-0.067132	-0.054329	-0.034629	-0.012333	0.009553
0.000618	-0.052631	-0.040382	-0.026392	-0.008904	0.010853
0.000820	-0.061422	-0.047123	-0.028338	-0.009079	0.009107
0.001072	-0.069756	-0.055724	-0.036793	-0.015227	0.008884
0.000655	-0.055363	-0.043477	-0.028286	-0.010781	0.008964
0.000873	-0.064044	-0.051530	-0.035942	-0.016808	0.005584
0.000875	-0.060043	-0.046270	-0.029765	-0.009939	0.013792
0.000860	-0.065346	-0.049600	-0.033415	-0.014309	0.007078
0.000721	-0.056667	-0.043761	-0.026372	-0.009678	0.008854

3.7f. Conclusions and applications.

The simulation results are quite encouraging. It appears that the bootstrap methodology can be a contribution to the stereological literature. This opens up a lot of new possibilities, in particular, we can do approximate inference. For example, an approximate confidence interval of $E(\frac{1}{Y})$ can be derived from $P(Q_\alpha < T_2 < Q_{1-\alpha}) = 1-2\alpha$ where we recall

$$T_2 = \frac{1}{n-1} \sum_{i=2}^{n} \frac{1}{Y_{n,i}} - E(\frac{1}{Y}) .$$

A naive way is to replace Q_α and $Q_{1-\alpha}$, the quantiles of T_2 by their bootstrap estimates.

3.8. Truncation.

If we observe only $y \geq y_0$, following the same argument as in section 2.5 we are back to the no truncation case by considering $\sqrt{Y^2-y_0^2}$ instead of Y and the distribution of $\sqrt{\theta^2-y_0^2}$ given $\theta \geq y_0$ instead of $G(\theta)$.

Suppose $G(\theta)$ has density $g(\theta)$ then

$$g(\theta|\theta \geq y_0) = \frac{g(\theta)}{\int_{y_0}^{\infty} g(\theta)d\theta} \propto g(\theta) .$$

Let

$$\xi = \sqrt{\theta^2-y_0^2}$$

$$\xi^2 = \theta^2 - y_0^2$$

$$2\xi d\xi = 2\theta d\theta$$

$$\frac{d\theta}{d\xi} = \frac{\xi}{\theta} = \frac{\xi}{\sqrt{\xi^2+y_0^2}} .$$

So

$$g_\xi(\xi|\theta \geq y_0) = \frac{\xi}{\sqrt{\xi^2+y_0^2}} \; g_\theta(\sqrt{\xi^2+y_0^2}\,|\theta \geq y_0)$$

One consequence is that

$$\int_0^{\infty} \frac{1}{\xi} g(\xi|\theta \geq y_0)d\xi < \infty$$

CHAPTER IV

THIN SLICE OF THICKNESS 2τ

4.1. The basic formulas.

(4.1.1)
$$1 - H(y|\uparrow) = \int_y^\infty \frac{\sqrt{\theta^2-y^2}+\tau}{\theta+\tau}\, dG(\theta|\uparrow)$$

(4.1.2)
$$dG(\theta|\uparrow) = \frac{(\theta+\tau)\, dG(\theta)}{\mu_G+\tau}$$

where $\mu_G = \int \theta\, dG(\theta)$. So

(4.1.3)
$$1 - H(y|\uparrow) = \int_y^\infty \frac{\sqrt{\theta^2-y^2}+\tau}{\mu_G+\tau}\, dG(\theta) \quad .$$

4.2. Decomposition and inversion.

We shall see in this section that the three cases a) G continuous, b) G mixed, c) G discrete have to be treated separately. This fact seems to be unnoticed; all the formulas derived thus far in the literature are valid only in the continuous case. The results of 4.2b, 4.2c, and 4.2d are entirely new to the best of our knowledge. Section 4.2a contains no new results. They can be found in Coleman (1979)

4.2a. If G is continuous, then so is $H(y|\uparrow)$ and we can differentiate (4.1.3) to get

(4.2.1)
$$h(y|\uparrow) = \frac{\tau\, g(y)}{\mu_G+\tau} + \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}}\, \frac{g(\theta)}{\mu_G+\tau}\, d\theta$$

which can be inverted to give

$$(4.2.2)\quad g(\theta) = -\sqrt{\frac{2}{\pi}}\ \frac{\mu_G+\tau}{\tau}\ \frac{d}{d\theta}\int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right) h(y|\uparrow)dy$$

where

$$f(\omega) = \sqrt{2\pi}\ e^{\frac{\omega^2}{2}}\ \{1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\omega} e^{-\frac{1}{2}u^2}\ du\}\ .$$

Integrating (4.2.2)

$$(4.2.3)\quad 1 - G(\theta) = \sqrt{\frac{2}{\pi}}\ \frac{\mu_G+\tau}{\tau}\int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right) h(y|\uparrow)dy\ .$$

Setting $\theta = 0$ yields μ_G as a function of $H(\cdot|\uparrow)$,

$$(4.2.4)\qquad 1 = \sqrt{\frac{2}{\pi}}\ \frac{\mu_G+\tau}{\tau}\int_0^\infty f\left(\frac{y\sqrt{2\pi}}{2\tau}\right) h(y|\uparrow)dy\ .$$

4.2b. <u>A basic fact</u>.

$H(y|\theta,\uparrow)$ has a discrete component at $y = \theta$, this can be explained physically, since the observed y is the profile (maximum) radius. If we let t be the perpendicular distance of the center of the sphere from the center of the slice, then for $0 \le t \le \tau$, the observed y is θ, the radius of the sphere. Since $H(y|\theta,\uparrow)$ has a discrete component at $y = \theta$, so if G put mass at $\theta = \theta_1,\theta_2,\ldots,\theta_m$ then $H(y|\uparrow) = \int H(y|\theta,\uparrow)\ dG(\theta|\uparrow)$ would also put mass at the same points $\theta_1,\ldots,\theta_m$.

Thus, if G is not continuous, we cannot differentiate (4.1.3) to get (4.2.1). We can see that the single equation (4.2.1) cannot be right since it would treat probability density and probability (the mass put on $\theta_1,\ldots,\theta_m$) on an equal footing.

The correct formulas are the following: Let $G = G_c + G_d$ where

G_c - continuous part of G

$G_d = \sum_{i=1}^{m} g_d(\theta_i)\delta_{\theta_i}$ - discrete part of G, m can be ∞.

4.2c. Assume G is mixed, i.e. $\int dG_c > 0$, $\int dG_d > 0$. We will treat the purely discrete case later. Then $H(\cdot|\uparrow) = H_c + H_d$ and

$$H_d = \frac{\tau\, G_d}{\mu_G+\tau} \tag{4.2.5}$$

Note: This implies H_d and G_d have the same jump points.

$$h_c(y) = \frac{\tau\, g_c(y)}{\mu_G+\tau} + \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}}\,\frac{g_c(\theta)}{\mu_G+\tau}\,d\theta + \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}}\,\frac{1}{\mu_G+\tau}\,dG_d(\theta) \tag{4.2.6}$$

(4.2.6) can be rewritten as

$$h_c(y) - \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}}\,\frac{1}{\tau}\,dH_d(\theta) = \frac{\tau\, g_c(y)}{\mu_G+\tau} + \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}}\,\frac{g_c(\theta)}{\mu_G+\tau}\,d\theta\ . \tag{4.2.7}$$

(4.2.7) has the same form as (4.2.1) with h in (4.2.1) replaced by the left hand side of (4.2.7), g in (4.2.1) is replaced by g_c. Thus (4.2.7) can be inverted to

$$g_c(\theta) = -\sqrt{\frac{2}{\pi}}\,\frac{\mu_G+\tau}{\tau}\,\frac{d}{d\theta}\left\{\int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right)\left[h_c(y) - \int_y^\infty \frac{y}{\sqrt{\xi^2-y^2}}\,\frac{1}{\tau}\,dH_d(\xi)\right]\cdot dy\right\}. \tag{4.2.8}$$

Integrating (4.2.8) gives

$$(4.2.9) \quad \int dG_c - G_c(\theta) = \sqrt{\frac{2}{\pi}}\ \frac{\mu_G+\tau}{\tau} \int_\theta^\infty f\left(\frac{\sqrt{2\pi(y^2-\theta^2)}}{2\tau}\right)\left[h_c(y) - \int_y^\infty \frac{y}{\sqrt{\xi^2-y^2}}\,\frac{1}{\tau}\, dG_d(\xi)\right] \cdot dy$$

Inverting (4.2.5) gives

$$(4.2.10) \qquad G_d(\theta) = \frac{\mu_G+\tau}{\tau} H_d(\theta) \ .$$

So, except for μ_G and $\int dG_c$, (4.2.9) and (4.2.10) give G_c and G_d in terms of H_c and H_d. Set $\theta = 0$ in (4.2.9), we then have

$$(4.2.11) \quad \int dG_c = \sqrt{\frac{2}{\pi}}\ \frac{\mu_G+\tau}{\tau} \int_0^\infty f\left(\frac{y\sqrt{2\pi}}{2\tau}\right)\left[h_c(y) - \int_y^\infty \frac{y}{\sqrt{\xi^2-y^2}}\,\frac{1}{\tau}\, dH_d(\xi)\right]dy \ .$$

Set $\theta = \infty$ in (4.2.10) and using the fact $\int dG_c + \int dG_d = 1$ we have

$$(4.2.12) \qquad 1 - \int dG_c = \frac{\mu_G+\tau}{\tau} \int dH_d \ .$$

(4.2.11) and (4.2.12) are two equations in the two unknowns $\int dG_c$ and μ_G, thus they can be solved to give $\int dG_c$ and μ_G in terms of H_c and H_d.

4.2d. G is discrete.

$$G = G_d = \sum_{i=1}^{m} g_i \, \delta_{\theta_i}$$

(4.2.5) and (4.2.6) become

(4.2.13) $$H_d = \frac{\tau}{\mu_G+\tau} G$$

(4.2.14) $$h_c(y) = \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}} \frac{1}{\mu_G+\tau} dG(\theta)$$

(4.2.13) $\Rightarrow$ $$\int H_d = \frac{\tau}{\mu_G+\tau} .$$

Since

$$\int H_d + \int H_c = 1 , \quad \int dH_c = \frac{\mu_G}{\mu_G+\tau} .$$

Normalizing (4.2.13) and (4.2.14) we get

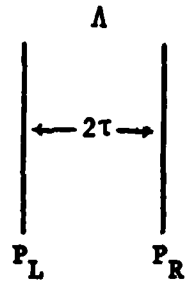
(4.2.15) $$\overline{H}_d = G$$

$$\overline{h}_c(y) = \frac{h_c(y)}{\int dH_c}$$

(4.2.16) $$= \int_y^\infty \frac{y}{\sqrt{\theta^2-y^2}} \frac{1}{\mu_G} dG(\theta) .$$

Observation 4.2.1: (4.2.16) is just (3.1.3) with h replaced by $\overline{h}_c$, so as far as $\overline{h}_c$ is concerned, we are back to the random plane case.

A physical explanation of observation 4.2.1: Let Λ denotes a thin slice of thickness 2τ, P_L and P_R the two plane sections at the two ends of Λ as shown in the diagram.



Then $\bar{h}_c(y)$ is just the conditional density of y given that a sphere is hit by the thin slice but the center of the sphere lies outside the interior of Λ (that is, either P_L hits the right half of the sphere or P_R hits the left half of the sphere). Thus $\bar{h}_c(y)$ is just the conditional density of y given that a sphere is hit by a random plane.

Note: For the mixed and continuous case, the above argument fails. We can no longer say that $\bar{h}_c(y)$ is the conditional density of y given that the left half of a sphere hits P_R or the right half of the sphere hits P_L. Because of the continuity of G_c, the discrete part of $H(\cdot|\theta,\dagger)$ (corresponds to when the center of the sphere lies in the interior of the thin slice Λ) can contribute to H_c, the continuous part of $H(\cdot|\dagger)$. This contribution is represented by the term $\dfrac{\tau\, g_c(y)}{\mu_G+\tau}$ in equation 4.2.6.

By observation 4.2.1, the inversion of (4.2.16) has already been done in section 3.1. Corresponding to (3.1.4), we have

$$(4.2.17) \qquad 1 - G(\theta) = \frac{\displaystyle\int_\theta^\infty \frac{\bar{h}_c(y)}{\sqrt{y^2-\theta^2}}\,dy}{\displaystyle\int_0^\infty \frac{1}{y}\,\bar{h}_c(y)dy} .$$

4.3. Decomposition of data, H_n and F_{Q*}.

In this section, we are going to introduce various decompositions that correspond to $H = H_c + H_d$.

4.3a. Notations concerning the data.

Given the data $y_1, y_2, \ldots, y_n$, define $\ell(n) = \#$ of distinct y's. Denote the distinct y's by $x_1, x_2, \ldots, x_{\ell(n)}$ with corresponding empirical frequencies $f_1, f_2, \ldots, f_{\ell(n)}$. Let

$$G = G_c + G_d$$

$$G_d = \sum_{i=1}^{m} g_i \delta_{\theta_i} \ , \quad \Theta = \{\theta_1, \theta_2, \ldots, \theta_m\} \ .$$

Define n_i = empirical frequency of θ_i in the data $y_1, y_2, \ldots, y_n$, $i = 1, \ldots, m$.

4.3b. Decomposition of data.

We are going to decompose the data according to whether $f_i > 1$ or not.

i. "Discrete" data (corresponds to $f_i > 1$). Let

$$n_d = \sum_{i=1}^{n} f_i \, I_{\{f_i > 1\}}$$

$$m(n) = \sum_{i=1}^{n} I_{\{f_i > 1\}} \ .$$

That is, n_d is the number of observations in the "discrete" data and $m(n)$ is the number of "distinct" values. Denote the distinct values by $\phi_1, \phi_2, \ldots, \phi_{m(n)}$. Define $\Phi_n = \{\phi_1, \phi_2, \ldots, \phi_{m(n)}\}$.

ii. "Continuous" data (corresponds to $f_i = 1$). $n_c = n - n_d$ is the number of observations in the "continuous" data. Denote them by $z_1, z_2, \ldots, z_{n_c}$.

Roughly speaking, we are regarding the "discrete" data as a sample from $\overline{H}_d$ and the "continuous" data as a sample from $\overline{H}_c$.

Note: When there is measurement error, exact ties may not be observed. We then need to consider all the values that cluster together as tied values. This would increase the probability of misclassification of data. Of course, the criteria of identifying a cluster depends on our knowledge of the size of the measurement error.

4.3c. <u>Decomposition of H_n</u>.

Let

$$H_n(y) = \frac{1}{n} \sum_{i=1}^{n} I\{y_i \le y\} = \frac{1}{n} \sum_{i=1}^{\ell(n)} f_i \, I\{x_i \le y\}$$

be the sample c.d.f.

Define

$$H_{nd}(y) = \frac{1}{n} \sum_{i=1}^{\ell(n)} f_i \, I\{x_i \le y\} \cdot I\{f_i > 1\} .$$

For continuous distribution, the probability of observing the same number twice is zero, so

(4.3.1) $\qquad \Phi_n \subset \Theta$ with probability 1 .

Therefore, an equivalent definition is

$$H_{nd}(y) = \frac{1}{n} \sum_{i=1}^{m} n_i \, I\{\theta_i \le y\} \, I\{n_i > 1\}$$

$$H_{nc}(y) = H_n(y) - H_{nd}(y)$$

$$= \frac{1}{n} \sum_{i=1}^{\ell(n)} I_{\{f_i = 1\}} I_{\{x_i \le y\}}$$

$$= \frac{1}{n} \sum_{i=1}^{n_c} I_{\{z_i \le y\}} .$$

H_c, H_d, H_{nc}, H_{nd} are all subdistribution functions, we can normalize them to distribution functions $\overline{H}_c$, $\overline{H}_d$, $\overline{H}_{nc}$, $\overline{H}_{nd}$. That is,

$$\overline{H}_c = H_c/\int dH_c$$

$$\overline{H}_d = H_d/\int dH_d$$

$$\overline{H}_{nc} = H_{nc}/\int dH_{nc}$$

$$= H_{nc}/(n_c/n)$$

$$= \frac{1}{n_c} \sum_{i=1}^{n_c} I_{\{z_i \le y\}} .$$

$$\overline{H}_{nd}(y) = H_{nd}/\int dH_{nd}$$

$$= H_{nd}/(n_d/n)$$

$$= \frac{1}{n_d} \sum_{i=1}^{\ell(n)} f_i \, I_{\{x_i \le y\}} \cdot I_{\{f_i > 1\}}$$

$$= \frac{1}{n_d} \sum_{i=1}^{m} n_i \, I_{\{\theta_i \le y\}} \cdot I_{\{n_i > 1\}} .$$

4.3d. Decomposition of F_{Q*}.

Recall the identification

$$H(y|\theta,\uparrow) \leftrightarrow F_\theta(y)$$

$$G(\theta|\uparrow) \leftrightarrow Q(\theta)$$

$$H(y|\uparrow) \leftrightarrow F_Q(y) \ .$$

Let

$$G = G_c + G_d \ , \quad G_d = \sum_{i=1}^{m} g_i \delta_{\theta_i}$$

$$Q = Q_c + Q_d \ .$$

By (4.1.2)

$$Q(\theta) = \frac{\int_0^\theta (\xi+\tau) dG(\xi)}{\mu_G+\tau}$$

and

$$Q_d = \sum_{i=1}^{m} \frac{\theta_i+\tau}{\mu_G+\tau} g_i \delta_{\theta_i}$$

$$= \sum_{i=1}^{m} q_i \delta_{\theta_i} \ .$$

Equation (4.2.5) can then be rewritten as

$$H_d = \sum_{i=1}^{m} \frac{\tau}{\tau+\theta_i} q_i \delta_{\theta_i} \ . \tag{4.3.2}$$

(4.3.2) motivates the following decomposition of F_{Q*},

$$F_{Q*} = F^{(n)}_{Q*,c} + F^{(n)}_{Q*,d}$$

where n is the sample size.

In this dissertation, the only time we use this decomposition is when $Q^* \varepsilon \mathcal{G}_n$, where $\mathcal{G}_n$ = the class of discrete distributions with weights at $x_1, x_2, \ldots, x_{\ell(n)}$.

The idea beind this is that Q^* is a candidate for being an estimate of Q. With probability 1, $\Phi_n \uparrow \Theta$ (see section 4.5) where we recall $\Phi_n = \{\phi_1, \phi_2, \ldots, \phi_{m(n)}\} = \{\theta_i : n_i > 1\}$ so that if Q^* puts mass q_i^* at θ_i, the corresponding estimate of Q_d is

$$\sum_{i=1}^{m} q_i^* I_{\{n_i > 1\}} \delta_{\theta_i} .$$

Based on (4.3.2), we define

$$(4.3.3) \qquad F^{(n)}_{Q*,d} = \sum_{i=1}^{m} \frac{\tau}{\tau+\theta_i} q_i^* I_{\{n_i > 1\}} \delta_{\theta_i}$$

and

$$(4.3.4) \qquad F^{(n)}_{Q*,c} = F_{Q*} - F^{(n)}_{Q*,d} .$$

4.4. Methods based on inversion formulas.

4.4a. G continuous.

Replace $h(y|\uparrow)dy$ in (4.2.3), (4.2.4) by $d\tilde{H}_n(y)$. If $\tilde{H}_n$ is the sample c.d.f., we are using classical back substitution; if $\tilde{H}_n$ is the sample c.d.f. smoothed by piecewise Lagrange interpolation, we are using Anderssen and Jakeman's approach.

4.4b. G mixed.

Replace dH_d by $d\tilde{H}_{nd}$ and $h_c(y)dy$ by $d\tilde{H}_{nc}(y)$ in equations (4.2.9), (4.2.10), (4.2.11) and (4.2.12), where $\tilde{H}_{nd}$ and $\tilde{H}_{nc}$ are estimates of H_d, H_c respectively. For example, we can let $\tilde{H}_{nd} = H_{nd}$ where H_{nd} is defined in section 4.3c and $\tilde{H}_{nc}$ is a smoothed version of H_{nc}.

4.4c. G discrete.

Here, we have a choice since we have two inversion formulas, namely (4.2.15) and (4.2.17).

We can obtain an estimate of G by simply replacing $\bar{H}_d$ by an estimate. $\tilde{\bar{H}}_{nd}$ in (4.2.15).

Alternatively, we can replace $\bar{h}_c(y)dy$ in (4.2.17) by $d\tilde{\bar{H}}_{nc}(y)$ where $\tilde{\bar{H}}_{nc}$ is some estimate of $\bar{H}_c$.

For example, we can use $\bar{H}_{nd}$ to estimate $\bar{H}_d$ and a smoothed version of $\bar{H}_{nc}$ to estimate H_c where $\bar{H}_{nc}$ and $\bar{H}_{nd}$ are defined in section 4.3c.

4.5. A closer look at the discrete case.

The stage is set in section 4.4c. We see from there that we can use either (4.2.15) or (4.2.17) to derive an estimate of G.

4.5a. Treating part of thin slice data as planar data.

We have more flexibility than is indicated in section 4.4c. (4.2.17) is just the inversion of (4.2.16) and observation 4.2.1 gives us a random plane interpretation of (4.2.16). Thus if we can separate from the data $y_1, y_2, \ldots, y_n$ the part which is a sample from $\overline{H}_c$, we can apply to that part of the data all the methods we have discussed in chapter III. A reasonable way to separate the data is to regard the "continuous" data defined in section 4.3b as a sample from $\overline{H}_c$. If m is finite, a justification for doing so is

$$(4.5.1) \qquad P(\text{misclassification of data}) \to 0 \quad \text{as} \quad n \to \infty .$$

We shall prove (4.5.1) later. Actually, more can be said in the case when m is finite.

4.5b. A basic question.

Returning to the point we made earlier, we can proceed from either (4.2.15) or (4.2.16) (or its inversion (4.2.17)) to obtain an estimate of G. But both equations contain information, thus we should make use of both of them instead of just one. Of course, we can always take a simple average or weighted average of two estimates, one based on (4.2.15) and the other on (4.2.16). But are there other ways, ways which are more built into the procedure for extracting the information contained in both equations? We

will see from section 4.7 that the minimum distance method is one such way. In the remainder of this section, we will discuss another procedure which uses both equations and which has its own appeal.

4.5c. Relationship between Φ_n and Θ .

We shall prove the following results for the case when G is mixed, thus including the discrete case as a special case. Let

$$G = G_c + G_d$$

$$G_d = \sum_{i=1}^{m} g_i \, \delta_{\theta_i} \, , \quad m \text{ can possibly equal infinity .}$$

Equation (4.2.5) tells us that G and H have the same points of jump, moreover if G put mass g_i at θ_i, then H put mass $h_i = \frac{\tau}{\tau+\theta_i} g_i$ at θ_i. By the strong law of large numbers, $\frac{n_i}{n} \to h_i$ with probability 1 so that for almost sample sequences, θ_i will be included in the "discrete" data if n is sufficiently large. In other words, if we let

$$\text{(4.5.2)} \qquad \Phi_\infty = \bigcup_{n=1}^{\infty} \Phi_n \, , \quad \text{note that } \Phi_n \subset \Phi_{n+1} \quad \text{(see section 4.3b)}$$

then

$$\text{(4.5.3)} \qquad P(\theta_i \in \Phi_\infty) = 1 \quad \text{for } i \text{ fixed.}$$

Since there can only be countably many points of jump

$$P\left(\bigcap_{i=1}^{m} (\theta_i \in \Phi_\infty)\right) = 1 ,$$

that is,

$$P(\Theta \subset \Phi_\infty) = 1$$

(4.3.1) $\Rightarrow$ $\Phi_\infty \subset \Theta$ with probability 1. Therefore $P\ (\Theta = \Phi_\infty) = 1$. Since we can never observe the whole sample sequence, it would be nice if we can say something about $P(\Phi_n = \Theta)$ for n large.

If $m < \infty$, (4.5.2) and (4.5.3) together imply $P(\theta_i \in \Phi_n) \to 1$ as $n \to \infty$. Since m is finite

(4.5.4) $$P\left(\bigcap_{i=1}^{m} (\theta_i \in \Phi_n)\right) \to 1 \quad \text{as} \quad n \to \infty .$$

Now

$$\bigcap_{i=1}^{m} (\theta_i \in \Phi_n) \Longleftrightarrow \Theta \subset \Phi_n$$

Also, $\Phi_n \subset \Theta$ with probability one. Therefore

(4.5.5) $$P(\Theta = \Phi_n) \to 1 \quad \text{as} \quad n \to \infty .$$

Note that since $\Phi_\infty \subset \Theta$ with probability one, the only way in which data can be misclassified is to have some θ_i not belonging to Φ_n. So

$$P(\text{misclassification of data})$$

$$= P\left(\bigcup_{i=1}^{m} (\theta_i \notin \Phi_n)\right)$$

$$= P\left(\left(\bigcap_{i=1}^{m} (\theta_i \in \Phi_n)\right)^c\right) \to 0 \quad \text{by (4.5.4)} .$$

Thus, we have proven (4.5.1).

If $m = \infty$, Φ_n being a finite set cannot equal Θ which is countable. So

$$P(\Phi_n = \Theta) = 0 \quad \text{for all } n .$$

However, we have the following result: Given $\varepsilon > 0$, there exists $\theta_{i_1}, \theta_{i_2}, \ldots, \theta_{i_{m_\varepsilon}}$ such that

$$\sum_{j=1}^{m_\varepsilon} g_{i_j} > 1-\varepsilon$$

$$\sum_{j=1}^{m_\varepsilon} q_{i_j} > 1-\varepsilon \quad .$$

Let

$$\Theta_\varepsilon = \{\theta_{i_1}, \theta_{i_2}, \ldots, \theta_{i_{m_\varepsilon}}\} \quad ,$$

then

$$P(\Phi_n \supset \Phi_\varepsilon) \to 1 \quad \text{as} \quad n \to \infty .$$

4.5d. <u>G discrete, m finite - further reduction</u>.

We already mentioned that we can regard the "continuous" data as a sample from $\overline{h}_c$, using (4.5.1) as justification. We also mention that we can then apply all the methods we discussed in chapter III to the "continuous" data. In sections 1.9 and 3.6, we see that if we know the support of G, then the computations of the MLE and minimum distance estimate are much simplified. Moreover, they give far better estimates than Anderssen and Jakeman's procedure which could not take advantage

AD-A120 687 A MIXING DISTRIBUTION APPROACH TO ESTIMATING PARTICLE SIZE DISTRIBUTIONS(U) STANFORD UNIV CA DEPT OF STATISTICS A Y KUK 19 OCT 82 TR-328 N00014-76-C-0475 2/2

UNCLASSIFIED F/G 12/1 NL

END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

the additional piece of information. While it is quite impossible to determine the support of G from random plane data, in our present case, (4.5.5) tells us that if n is sufficiently large, then with large probability, the support of G can be determined from the data. In other words, thin slice data contain more information than planar data and the additional information can be put to good use. Essentially, we are using (4.2.15) to determine $\theta_1,\theta_2,\ldots,\theta_m$, the support of $G = \sum_{i=1}^{m} g_i \delta_{\theta_i}$ and (4.2.16) to determine $g_1,g_2,\ldots,g_m$.

4.6. Miscellaneous lemmas.

In this section we state a number of lemmas which will be used in section 4.7, their proofs can be found in A.3 of the Appendix.

Lemma 4.6.1. With probability one, $\|H_n-H\| \to 0$.

Lemma 4.6.2. With probability one, $\|H_{nd}-H_d\| \to 0$.

Lemma 4.6.3. With probability one, $\|H_{nd}-H_c\| \to 0$.

Lemma 4.6.4. For almost every sample sequence, there exists a $Q_n^* \in \mathcal{G}_n$ for each n such that

$$Q_n^* \xrightarrow{\mathcal{D}} Q$$

and

$$Q_n^*(\theta_i) - Q_n^*(\theta_i-) = Q(\theta_i) - Q(\theta_i-)$$

for n sufficiently large.

Note: See Section 4.3d for the definition of $\mathcal{G}_n$.

Lemma 4.6.5. For almost every sample sequence, let Q_n^* be as defined in lemma 4.6.4, $\|F_{Q_n^*}-F_Q\| \to 0$.

Note that $Q = G(\cdot|\dagger)$, $F_Q = H(\cdot|\dagger)$ and we are in the thin slice case.

Lemma 4.6.6. For almost every sample sequence, let Q_n^* be as defined in lemma 4.6.4, $\|F_{Q_n^*,d}^{(n)} - H_d\| \to 0$.

Lemma 4.6.7. For almost every sample sequence, let Q_n^* be as defined in lemma 4.6.4, $\|F_{Q_n^*,c} - H_c\| \to 0$.

Lemma 4.6.8. If $F_{Q_n} \to F_Q$, then $Q_n \to Q$.

Lemma 4.6.9. If G is discrete, then for almost every sample sequence, there exists Q_n^*, Q_n^* discrete, puts weights at $\phi_1, \phi_2, \ldots, \phi_{m(n)}$ such that the statements of lemmas 4.6.4, 4.6.5, 4.6.6 and 4.6.7 still hold.

4.7. Minimum distance method.

4.7a. Description.

Procedure 1. $\hat{Q}_n$ minimizes $\|F_Q - H_n\|$ among all $Q \in \mathcal{Q}_n$.

This procedure is a linear programming problem with $4\ell(n)+1$ equations excluding the nonnegativity constraints and $\ell(n)+1$ variables.

Procedure 2. Among all $Q \in \mathcal{Q}_n$, Q_n minimizes

$$\|F_{Q,c}^{(n)} - H_{nc}\| + \|F_{Q,d}^{(n)} - H_{nd}\|$$

where $F_{Q,c}^{(n)} + F_{Q,d}^{(n)} = F_Q$ is a decomposition of F_Q (see section 4.3d). This is a linear programming problem with $4n_c + 2m(n) + 1$ linear constraints and $\ell(n)+2$ variables.

Procedure 3. Among all $Q \in \mathcal{Q}_n$, $\hat{Q}_n$ minimizes

$$d(F_Q, H_n) = \int (F_{Q,c}^{(n)}(y) - H_{nc}(y))^2 \, dH_{nc}(y)$$

$$+ \int \{(F_{Q,d}^{(n)}(y) - F_{Q,d}^{(n)}(y-)) - (H_{nd}(y) - H_{nd}(y-))\}^2 \, dH_{nd}(y) .$$

Note: The second integral can be rewritten as

$$\sum_{i=1}^{m} \frac{n_i}{n} \left(\frac{\tau}{\tau+\theta_i} q_i - \frac{n_i}{n}\right)^2 I_{\{n_i > 1\}}$$

where $q_i \geq 0$ is the weight Q put at θ_i.

This is a quadratic programming problem.

4.7b. Consistency of $\hat{Q}_n$.

For procedure 1. For almost every sample sequence, let Q_n^* be as defined in lemma 4.6.4. By lemma 4.6.5 $\|F_{Q_n^*} - F_Q\| \to 0$. By lemma 4.6.1 $\|H_n - F_Q\| \to 0$. Combining, $\|F_{Q_n^*} - H_n\| \to 0$. By definition,

$$\|F_{\hat{Q}_n} - H_n\| \leq \|F_{Q_n^*} - H_n\| \to 0 .$$

This implies $\|F_{\hat{Q}_n} - F_Q\| \to 0$. By lemma 4.6.8 $\hat{Q}_n \to Q$.

For procedure 2. For almost every sample sequence, let Q_n^* be as defined in lemma 4.6.4. By lemmas 4.6.6 and 4.6.7

$$\|F^{(n)}_{Q_n^*,c} - H_c\| \to 0$$

$$\|F^{(n)}_{Q_n^*,d} - H_d\| \to 0 .$$

These together with lemmas 4.6.2 and 4.6.3 imply

$$\|F^{(n)}_{Q_n^*,c} - H_{nc}\| + \|F^{(n)}_{Q_n^*,d} - H_{nd}\| \to 0$$

from which we conclude

$$(4.7.1) \qquad \|F^{(n)}_{\hat{Q}_n,c} - H_{nc}\| + \|F^{(n)}_{\hat{Q}_n,d} - H_{nd}\| \to 0$$

by definition of $\hat{Q}_n$.

(4.7.1) together with lemmas 4.6.2 and 4.6.3 imply

$$(4.7.2) \qquad \|F^{(n)}_{\hat{Q}_n,c} - H_c\| + \|F^{(n)}_{\hat{Q}_n,d} - H_d\| \to 0 \ .$$

Since

$$H = H_c + H_d = F_Q$$

$$F_{\hat{Q}_n} = F^{(n)}_{\hat{Q}_n,c} + F^{(n)}_{\hat{Q}_n,d}$$

$$(4.7.2) \quad \Rightarrow \quad \|F_{\hat{Q}_n} - F_Q\| \to 0 \ .$$

So by lemma 4.6.8 $\hat{Q}_n \xrightarrow{\mathcal{L}} Q$.

<u>For procedure 3</u>. We assume $m < \infty$ in addition. For almost every sample sequence, let Q^*_n be defined as in lemma 4.6.4

$$(4.7.3) \quad \int (F^{(n)}_{Q^*_n,c}(y) - H_{nc}(y))^2 \, dH_{nc}(y)$$

$$\leq \|F^{(n)}_{Q^*_n,c} - H_{nc}\|^2 \to 0 \text{ by lemma 4.6.3 and lemma 4.6.7.}$$

Let Q^*_n put mass q^*_i at θ_i. By construction, $q^*_i = q_i$ for n sufficiently large (lemma 4.6.4), and $\frac{n_i}{n} \to h_i = \frac{\tau}{\tau+\theta_i} q_i$ by the strong

law of large numbers. Therefore, since m is finite

$$\sum_{i=1}^{m} \frac{n_i}{n} \left(\frac{\tau}{\tau+\theta_i} q_i^* - \frac{n_i}{n}\right)^2 I_{\{n_i > 1\}} \tag{4.7.4}$$

$$+ \sum_{i=1}^{m} h_i (h_i - h_i)^2 = 0 \ .$$

(4.7.3) and (4.7.4) together imply

$$d(F_{Q_n^*}, H_n) \rightarrow 0 \ .$$

By definition of $\hat{Q}_n$,

$$d(F_{\hat{Q}_n}, H_n) \rightarrow 0 \ .$$

That is

$$\int (F_{\hat{Q}_n, c}^{(n)}(y) - H_{nc}(y))^2 \, dH_{nc}(y) \rightarrow 0 \tag{4.7.5}$$

and

$$\sum_{i=1}^{m} \frac{n_i}{n} \left(\frac{\tau}{\tau+\theta_i} \hat{q}_i - \frac{n_i}{n}\right)^2 I_{\{n_i > 1\}} \rightarrow 0 \tag{4.7.6}$$

where $\hat{q}_i \geq 0$ is the mass $\hat{Q}_n$ puts at θ_i .

Making use of the lemma on P. 453, Choi and Bulgren (1968), it follows from (4.7.5) that

$$\| F_{\hat{Q}_n, c}^{(n)} - H_c \| \rightarrow 0 \ . \tag{4.7.7}$$

It follows from (4.7.6)

$$\hat{q}_i \to q_i \quad \text{since} \quad \frac{n_i}{n} \to h_i = \frac{\tau}{\tau+\theta_i} q_i \ .$$

Therefore

$$\hat{h}_i = \frac{\tau}{\tau+\theta_i} \hat{q}_i \to \frac{\tau}{\tau+\theta_i} q_i = h_i \ .$$

$$H_d = \sum_{i=1}^{m} h_i \delta_{\theta_i}$$

$$F^{(n)}_{\hat{Q}_n,d} = \sum_{i=1}^{m} \hat{h}_i \, I_{\{n_i > 1\}} \delta_{\theta_i} \ .$$

Since m is finite

$$\| F^{(n)}_{\hat{Q}_n,d} - H_d \| \to 0 \ .$$

It follows from this and (4.7.7) that

$$\| F_{\hat{Q}_n} - H \| \to 0 \ .$$

By lemma 4.6.8, $\hat{Q}_n \xrightarrow{\mathcal{D}} Q$.

4.7e. Consistency of $\hat{G}_n$.

We have already proved the consistency of $\hat{Q}_n = \hat{G}_n(\cdot|\uparrow)$ in section 4.7b. From (4.1.1)

$$G(\theta) = \frac{\int_0^{\theta} \frac{1}{\xi+\tau} \, dG(\xi|\uparrow)}{\int_0^{\infty} \frac{1}{\xi+\tau} \, dG(\xi|\uparrow)} \ .$$

Since $\frac{1}{\xi+\tau}$ as a function of ξ is continuous and bounded. By Theorem 4.4.2,

P. 89, Chung $\hat{G}_n(\cdot|+) \longrightarrow G(\cdot|+)$ implies

$$\hat{G}_n(\theta) = \frac{\int_0^\theta \frac{1}{\xi+\tau}\, d\hat{G}_n(\xi|+)}{\int_0^\infty \frac{1}{\xi+\tau}\, d\hat{G}_n(\xi|+)} \longrightarrow G(\theta)$$

for every θ.

So $\hat{G}_n \xrightarrow{\mathscr{D}} G$.

4.7d. <u>Discrete case - further simplification</u>.

If G is discrete, Procedures 1, 2 and 3 described in section 4.7a can be simplified as follows: we minimize among the class of discrete distributions that put weights at $\phi_1, \phi_2, \ldots, \phi_{m(n)}$ instead of among the class of discrete distributions that put weights at $x_1, x_2, \ldots, x_{\ell(n)}$. $m(n)$ can be considerably less than $\ell(n)$, so we can have considerable reduction in computation.

Consistency of $\hat{Q}_n$ and $\hat{G}_n$ can also be proven, the proof of section 4.7b is still valid. The role of lemmas 4.6.4, 4.6.5, 4.6.6, 4.6.7 is now played by lemma 4.6.9.

4.8. <u>Simulation results</u>.

G is discrete

$$\Theta = \{1,2,3,4,5\} \qquad g_i = 0.2 \qquad i = 1,\ldots,5 .$$

So

$$\mu_G = 3$$

$$\tau = 0.25 .$$

By (4.2.4)

$$h_i = \frac{\tau}{\tau+\mu_G} g_i$$

$$= \frac{0.25}{0.25+3} 0.2$$

$$= \frac{0.2}{13} .$$

Sample size $n = 500$.

We compare eight different methods, an estimate is computed only when $n_i > 1$, $i = 1,\ldots,5$, i.e., when we can successfully determine θ from the data.

The eight methods are described below:

1) By (4.2.6) $g_i \propto h_i$. So estimate g_i by

$$\hat{g}_i = \frac{n_i}{\sum_{i=1}^{5} n_i} , \; i = 1,\ldots,5 .$$

2) Consider the "continuous" data as planar data, use MLE with known support.

3) Consider the "continuous" data as planar data, use Choi and Bulgren's approach with known support.

4) Average of 1) and 2).

5) Average of 1) and 3).

6) Weighted average of 1) and 2). The weights being $\frac{n_d}{n}$ and $\frac{n_c}{n}$.

7) Weighted average of 1) and 3).

8) Procedure 3 of section 4.7. We use the simplified version (see Section 4.7d).

The results are as follow: In 94 out of the 100 trials, θ can be determined from the data, i.e., $n_i > 1$, $i = 1,\ldots,5$.

TABLE 21

A comparison of the 8 methods of estimating the probability mass function $P(\theta)$ of the size distribution $G(\theta)$ which puts mass 0.2 at $\theta = 1,2,3,4,5$ in terms of m.s.e. and the average of the estimates over 94 samples of size 100.

θ	1.00000	2.00000	3.00000	4.00000	5.00000
$P(\theta)$	0.20000	0.20000	0.20000	0.20000	0.20000
METHOD 1	0.21215	0.19677	0.19859	0.19768	0.19481
M.S.E.	0.00393	0.00298	0.00509	0.00354	0.00441
METHOD 2	0.19588	0.19857	0.20474	0.20010	0.20070
M.S.E.	0.00151	0.00101	0.00084	0.00065	0.00041
METHOD 3	0.20304	0.19455	0.20629	0.19864	0.19748
M.S.E.	0.00180	0.00145	0.00111	0.00093	0.00056
METHOD 4	0.20401	0.19767	0.20166	0.19889	0.19776
M.S.E.	0.00135	0.00105	0.00131	0.00111	0.00132
METHOD 5	0.20759	0.19566	0.20244	0.19816	0.19615
M.S.E.	0.00143	0.00113	0.00146	0.00120	0.00136
METHOD 6	0.19717	0.19843	0.20411	0.19998	0.20031
M.S.E.	0.00130	0.00089	0.00070	0.00059	0.00041
METHOD 7	0.20379	0.19471	0.20553	0.19863	0.19734
M.S.E.	0.00155	0.00126	0.00094	0.00084	0.00054
METHOD 8	0.20406	0.19341	0.20608	0.19622	0.20023
M.S.E.	0.00172	0.00144	0.00108	0.00111	0.00099

Comment: Method 1, which uses just the "discrete" data is clearly inferior to methods 2 and 3 which regard the "continuous" data as planar data. Methods 2 and 3 are themselves comparable. Method 8 seems to perform better than method 1 but not as good as methods 2 and 3.

Estimate 4. The simple average of estimates 1 and 2 is no better (maybe even worse) than estimate 2. Similarly, estimate 5, the simple average of estimates 1 and 3 is no better than estimate 3. However, if we take the weighted average, there is considerable improvement, giving us the best results. (Compare method 6 with method 2, method 7 with method 3).

Treating the "continuous" data as planar data, we are back to the random plane case, so a comparison between the performance of the procedures we proposed with the classical procedures is already given in section 3.6.

Recall also that in cases 1 and 2 of section 3.6, we assume we know θ. While this assumption is quite unrealistic in the random plane case, in our present thin slice case, in 94 out of the 100 trials, θ can actually be determined from the "discrete" data.

APPENDIX

A.1. In section 3.3, we give four conditions (c1-c4) which characterize $h(\cdot|\dagger)$ in the random plane case. The sample c.d.f. H_n does not belong to the admissible range $\{H(\cdot|\dagger)\}$. Since the Anderssen and Jakeman product integration estimate $\tilde{H}_n$ is just a smoothed version of H_n, there is no reason to believe that $\tilde{H}_n$ will belong to $\{H(\cdot|\dagger)\}$. To illustrate this, we will find a G such that $P(\tilde{h}_n$ violates c4) $> P$ for all n sufficiently large, and for some $P > 0$.

Consider $y_1, y_2, \ldots, y_n$, i.i.d. from the density

$$h(y|\dagger) = \int_y^\infty \frac{y}{\sqrt{\theta^2 - y^2}} \frac{1}{\mu_G} dG(\theta) ,$$

arrange them in increasing order as $y_{n,1} < y_{n,2} < \cdots < y_{n,n}$, define $y_{n,0} = 0$. We will consider piecewise linear interpolation, the simplest kind of product integration estimate. That is,

$$\tilde{h}_n(y) = \frac{\frac{1}{n}}{y_{n,i} - y_{n,i-1}} , \ y_{n,i-1} \leq y < y_{n,i} .$$

As it stands now, $\tilde{h}_n$ violates c1 of the characterization of section 3.3. However, this problem can be fixed easily, for example, by applying quadratic interpolation to points adjacent to the origin. But if c4 is violated which is the same as saying that $\tilde{G}_n$ is not a nondecreasing function, then there is no getting around. We will find a G such that $P(\tilde{h}_n$ violates c4) $> P$ for all n sufficiently large and for some $P > 0$.

Let

$$\theta_{n,1} = y_{n,n-2}$$

$$\theta_{n,2} = y_{n,n-1} \qquad \varepsilon_{n,1} = \theta_{n,2} - \theta_{n,1}$$

$$\theta_{n,3} = y_{n,n} \qquad \varepsilon_{n,2} = \theta_{n,3} - \theta_{n,2} .$$

Define

$$I_n = \frac{1}{\varepsilon_{n,1}} \int_{\theta_{n,1}}^{\theta_{n,2}} \frac{1}{\sqrt{y^2-\theta_{n,1}^2}}\, dy$$

$$J_n = \frac{1}{\varepsilon_{n,2}} \int_{\theta_{n,2}}^{\theta_{n,3}} \frac{1}{\sqrt{y^2-\theta_{n,1}^2}}\, dy$$

$$K_n = \frac{1}{\varepsilon_{n,2}} \int_{\theta_{n,2}}^{\theta_{n,3}} \frac{1}{\sqrt{y^2-\theta_{n,2}^2}}\, dy ,$$

so that

$$I_n + J_n < K_n \Rightarrow \tilde{h}_n \text{ violates c4}$$

$$\therefore\ P(\tilde{h}_n \text{ violates c4}) > P(I_n + J_n < K_n)$$

$$K_n - J_n = \frac{1}{\varepsilon_{n,2}} \int_{\theta_{n,2}}^{\theta_{n,3}} \frac{1}{\sqrt{y^2-\theta_{n,2}^2}} - \frac{1}{\sqrt{y^2-\theta_{n,1}^2}}\, dy .$$

Fact 1 : $\displaystyle \frac{1}{\sqrt{a-\delta}} - \frac{1}{\sqrt{a}} = \frac{\sqrt{a} - \sqrt{a-\delta}}{\sqrt{a}\,\sqrt{a-\delta}} > \frac{\delta}{2a\sqrt{a-\delta}} .$

Proof. If

$$f(\delta) = \sqrt{a-\delta}$$

$$f'(\delta) = \frac{-1}{2\sqrt{a-\delta}}$$

$$f''(\delta) = \frac{-1}{4(a-\delta)^{3/2}}$$

$$\Rightarrow \quad \sqrt{a-\delta} < \sqrt{a} - \frac{\delta}{2\sqrt{a}}$$

Since $\theta_{n,2} = \theta_{n,1} + \varepsilon_{n,1}$

$$y^2 - \theta^2_{n,2} = y^2 - \theta^2_{n,1} - 2\theta_{n,1}\varepsilon_{n,1} - \varepsilon^2_{n,1} .$$

Apply fact 1 with

$$a = y^2 - \theta^2_{n,1} \ , \ \delta = 2\theta_{n,1}\varepsilon_{n,1} + \varepsilon^2_{n,1} = \theta^2_{n,2} - \theta^2_{n,1} ,$$

$$K_n - J_n > \frac{1}{\varepsilon_{n,2}} \int_{\theta_{n,2}}^{\theta_{n,3}} \frac{\theta^2_{n,2} - \theta^2_{n,1}}{2(y^2-\theta^2_{n,1})} \ \frac{1}{\sqrt{y^2-\theta^2_{n,2}}} \, dy$$

$$> \frac{1}{\varepsilon_{n,2}} \int_{\theta_{n,2}}^{\theta_{n,3}} \frac{\theta^2_{n,2} - \theta^2_{n,1}}{2(\theta^2_{n,3} - \theta^2_{n,1})} \ \frac{1}{\sqrt{y^2-\theta^2_{n,2}}} \, dy .$$

Thus $K_n > I_n + J_n$ if

$$\text{(1)} \qquad \frac{1}{2\varepsilon_{n,2}} \ \frac{\theta^2_{n,2} - \theta^2_{n,1}}{\theta^2_{n,3} - \theta^2_{n,1}} \int_{\theta_{n,2}}^{\theta_{n,3}} \frac{1}{\sqrt{y^2-\theta^2_{n,2}}} \, dy > \frac{1}{\varepsilon_{n,1}} \int_{\theta_{n,1}}^{\theta_{n,2}} \frac{1}{\sqrt{y^2-\theta^2_{n-1}}} \, dy$$

Next,

$$\int \frac{du}{\sqrt{u^2-a^2}} = \log(u + \sqrt{u^2-a^2})$$

so that

$$\int_a^v \frac{du}{\sqrt{u^2-a^2}} = \log(v + \sqrt{v^2-a^2}) - \log a$$

Thus (1) becomes

$$(2) \qquad \frac{1}{2\epsilon_{n,2}} \frac{\theta_{n,2}^2 - \theta_{n,1}^2}{\theta_{n,3}^2 - \theta_{n,1}^2} (\log(\theta_{n,3} + \sqrt{\theta_{n,3}^2 - \theta_{n,2}^2}) - \log \theta_{n,2})$$

$$> \frac{1}{\epsilon_{n,1}} (\log(\theta_{n,2} + \sqrt{\theta_{n,2}^2 - \theta_{n,1}^2}) - \log \theta_{n,1}) .$$

Let

$$f(x) = \log(a+x) , \qquad x > 0$$

$$f'(x) = \frac{1}{(a+x)}$$

$$f''(x) = \frac{-1}{(a+x)^2}$$

$$f'''(x) = \frac{2}{(a+x)^3} .$$

We have the following

Fact 2: $\log(a+x) < \log a + \frac{1}{a} x$.

Fact 3: $\log(a+x) > \log a + \frac{1}{a} x - \frac{1}{2a^2} x^2$.

Apply fact 3 to the left hand side of (2) and fact 2 to the right hand side of (2), we see that (2) is satisfied if

$$(3) \qquad \frac{1}{2\epsilon_{n,2}^2} \frac{\theta_{n,2}^2 - \theta_{n,1}^2}{\theta_{n,3}^2 - \theta_{n,1}^2} (\log \theta_{n,3} + \frac{\sqrt{\theta_{n,3}^2 - \theta_{n,2}^2}}{\theta_{n,3}} - \frac{1}{2\theta_{n,3}^2} (\theta_{n,3}^2 - \theta_{n,2}^2) - \log \theta_{n,2})$$

$$> \frac{1}{\epsilon_{n,1}} (\log \theta_{n,2} + \frac{\sqrt{\theta_{n,2}^2 - \theta_{n,1}^2}}{\theta_{n,2}} - \log \theta_{n,1}) .$$

Apply fact 3 to $\log \theta_{n,3} = \log(\theta_{n,2} + \varepsilon_{n,2})$ on the left hand side of (3) and fact 2 to $\log \theta_{n,2} = \log(\theta_{n,1} + \varepsilon_{n,1})$ on the right hand side of (3). (3) is satisfied if

$$(4) \qquad \frac{1}{2\varepsilon_{n,2}} \frac{\theta^2_{n,2} - \theta^2_{n,1}}{\theta^2_{n,3} - \theta^2_{n,1}} \left(\frac{\varepsilon_{n,2}}{\theta_{n,2}} - \frac{1}{2\theta^2_{n,2}} \varepsilon^2_{n,2} + \frac{\sqrt{\theta^2_{n,3} - \theta^2_{n,2}}}{\theta_{n,3}} - \frac{1}{2\theta^2_{n,3}} (\theta^2_{n,3} - \theta^2_{n,2})\right)$$

$$> \frac{1}{\varepsilon_{n,1}} \left(\frac{\varepsilon_{n,1}}{\theta_{n,1}} + \frac{\sqrt{\theta^2_{n,2} - \theta^2_{n,1}}}{\theta_{n,2}}\right) .$$

Let $G(\theta)$ have support on $[0,1]$ so that

$$h(y|\uparrow) = \int_y^\infty \frac{1}{\sqrt{\theta^2 - y^2}} \frac{1}{\mu_G} dG(\theta)$$

also has support on $[0,1]$. Then with probability one

$$\theta_{n,1} \to 1 , \quad \theta_{n,2} \to 1 , \quad \theta_{n,3} \to 1$$

$$\varepsilon_{n,1} \to 0 , \quad \varepsilon_{n,2} \to 0 .$$

Note that

$$\theta^2_{n,3} - \theta^2_{n,2} = (\theta_{n,3} + \theta_{n,2})(\theta_{n,3} - \theta_{n,2})$$

$$\theta^2_{n,2} - \theta^2_{n,1} = (\theta_{n,2} + \theta_{n,1})(\theta_{n,2} - \theta_{n,1}) .$$

Taking the limit in (4), we have

(5) $$\frac{1}{2\varepsilon_{n,2}} \frac{\varepsilon_{n,1}}{\varepsilon_{n,1}+\varepsilon_{n,2}} (\varepsilon_{n,2} - \frac{\varepsilon_{n,2}^2}{2} + \sqrt{2\varepsilon_{n,2}} - \frac{1}{2}\varepsilon_{n,2}\cdot 2)$$

$$> \frac{1}{\varepsilon_{n,1}} (\varepsilon_{n,1} + \sqrt{2\varepsilon_{n,1}}) .$$

Since $\varepsilon_{n,1} \to 0$, $\varepsilon_{n,2} \to 0$ with probability one, $\sqrt{2\varepsilon_{n,2}}$ dominates the bracketed terms on the left hand side and $\sqrt{2\varepsilon_{n,1}}$ on the right hand side. So, we can look at

$$\frac{1}{2\varepsilon_{n,2}} \frac{\varepsilon_{n,1}}{\varepsilon_{n,1}+\varepsilon_{n,1}} \sqrt{\varepsilon_{n,2}} > \frac{1}{\varepsilon_{n,1}} \sqrt{\varepsilon_{n,1}}$$

$$\frac{\varepsilon_{n,1}}{\varepsilon_{n,1}+\varepsilon_{n,2}} > 2\sqrt{\frac{\varepsilon_{n,2}}{\varepsilon_{n,1}}}$$

(6) $$1 + \frac{\varepsilon_{n,2}}{\varepsilon_{n,1}} < \frac{1}{2}\sqrt{\frac{\varepsilon_{n,1}}{\varepsilon_{n,2}}} .$$

(6) is satisfied if $\frac{\varepsilon_{n,1}}{\varepsilon_{n,2}} > c$ for c sufficiently large, actually $c = 9$ will do. So $P(\tilde{h}_n$ violates c4)

$$> P(I_n + J_n < K_n)$$

$$> P(\frac{\varepsilon_{n,1}}{\varepsilon_{n,2}} > c) \quad \text{if } G \text{ has support on } [0,1] .$$

Now, if $y_1, y_2, \ldots, y_n$ are i.i.d. according to the uniform distribution on $[0,1]$, then

$$\frac{Y_{n,n-1}-Y_{n,n-2}}{Y_{n,n}-Y_{n,n-1}} \Bigg| \; Y_{n,n-2} = y_{n,n-2} \;\; \underset{=}{\mathcal{D}} \;\; \frac{x_{(1)}}{x_{(2)}-x_{(1)}}$$

where x_1, x_2 i.i.d. $\sim U(0,1)$

$$\Rightarrow \frac{Y_{n,n-1}-Y_{n,n-2}}{Y_{n,n}-Y_{n,n-1}} \;\; \underset{=}{\mathcal{D}} \;\; \frac{x_{(1)}}{x_{(2)}-x_{(1)}} .$$

Then

$$P\left(\frac{\varepsilon_{n,1}}{\varepsilon_{n,2}} > c\right) = P\left(\frac{x_{(1)}}{x_{(2)}-x_{(1)}} > c\right) = P > 0 .$$

But the uniform density on $[0,1]$ violates c1 and hence does not correspond to any G.

However, take

$$g(\theta|+) = \frac{c\theta^2}{\sqrt{1-\theta^2}} \qquad 0 < \theta < 1 .$$

Then

$$h(y|+) = \int_y^1 \frac{y}{\theta\sqrt{\theta^2-y^2}} \frac{c\theta^2}{\sqrt{1-\theta^2}} \, d\theta$$

$$= \frac{cy}{2} \int_y^1 \frac{2\theta \, d\theta}{\sqrt{\theta^2-y^2}\sqrt{1-\theta^2}}$$

let $\xi = \theta^2$, $d\xi = 2\theta \, d\theta$

$$h(y|\dagger) = \frac{cy}{2}\int_{y^2}^{1} \frac{d\xi}{\sqrt{\xi-y^2}\sqrt{1-\xi}}$$

$$= \frac{cy}{2}\, 2\sin^{-1}\left(\sqrt{\frac{\xi-y^2}{1-y^2}}\right)\Bigg|_{\xi=y^2}^{1}$$

$$= cy(\sin^{-1}(1) - \sin^{-1}(0))$$

$$= \frac{c\pi}{2}\, y \qquad \text{for} \quad 0 < y < 1 \ .$$

Thus $h(1-|\dagger) = \frac{c\pi}{2}$.

Then if $Y_1, Y_2, \ldots, Y_n$ i.i.d. according to the above $H(\cdot|\dagger)$

$$\frac{\epsilon_{n,1}}{\epsilon_{n,2}} = \frac{Y_{n,n-1}-Y_{n,n-2}}{Y_{n,n}-Y_{n,n-1}} \overset{\mathcal{D}}{=} \frac{H^{-1}(U_{n,n-1}) - H^{-1}(U_{n,n-2})}{H^{-1}(U_{n,n}) - H^{-1}(U_{n,n-1})}$$

where $u_1, u_2, \ldots, u_n$ i.i.d. $\sim U(0,1)$ and $u_{n,1} < \cdots < u_{n,n}$ their ordered value. But

$$\frac{H^{-1}(U_{n,n-1})-H^{-1}(U_{n,n-2})}{H^{-1}(U_{n,n})-H^{-1}(U_{n,n-1})}$$

$$= \frac{U_{n,n-1}-U_{n,n-2}}{U_{n,n}-U_{n,n-1}} \ \frac{h(W_{n,2}|\dagger)}{h(W_{n,1}|\dagger)}$$

where

$$U_{n,n-2} < W_{n,1} < U_{n,n-1}$$
$$U_{n,n-1} < W_{n,2} < U_{n,n} \ .$$

Since $W_{n,1} \to 1$, $W_{n,2} \to 1$ with probability one and $h(1-|\dagger)$ is finite

$$\frac{h(W_{n,2}|\dagger)}{h(W_{n,1}|\dagger)} \to 1 \qquad \text{with probability 1}$$

$$\therefore \lim_n P\left(\frac{\varepsilon_{n,1}}{\varepsilon_{n,2}} > c\right) = \lim_n P\left(\frac{U_{n,n-1}-U_{n,n-2}}{U_{n,n}-U_{n,n-1}} > c\right)$$

$$= P > 0 .$$

Since

$$P(\tilde{h}_n \text{ violates c4})$$

$$> P(I_n + J_n < K_n)$$

$$> P\left(\frac{\varepsilon_{n,1}}{\varepsilon_{n,2}} > c\right) \to P$$

$$P(\tilde{h}_n \text{ violates c4}) > P \quad \text{for } n \text{ sufficiently large.}$$

A.2. Proof of lemmas 3.5.1 and 3.5.2.

We first prove the following lemma.

Lemma: Let

$$Y_1, Y_2, \ldots, Y_n \cdots \quad \text{i.i.d.} \sim h_0(y|\dagger) = \int_y^\infty \frac{y}{\theta\sqrt{\theta^2 - y^2}} \, dG_0(\theta|\dagger).$$

Let $A = \inf\{\theta\colon G(\theta) = 1\}$, possibly ∞ then almost every sample sequence $\{y_k\}$ is dense in $(0,A)$.

Proof. The definition of $h_0(y|\dagger)$ and A together imply

(1) $$h_0(y|\dagger) > 0 \quad \text{for} \quad y \in (0,A) .$$

Let Q = the set of positive rationals

$$Q^* = Q \cap (0,A)$$

$$\mathcal{J} = \{(q-r, q+r),\ q \in Q^*,\ r \in Q\} ,$$

let $(a,b) \in \mathcal{J}$

(1) $\Rightarrow P(a,b) > 0$

$\Rightarrow$ for almost every sample sequence $\{y_k\}$

$\{y_k\} \cap (a,b) \neq \phi$.

Since $\mathcal{J}$ is countable, we conclude that for almost every sample sequence $\{y_k\}$.

$$\{y_k\} \cap I \neq \phi \qquad \forall\, I \in \mathcal{J} .$$

Since Q^* is dense in $(0,A)$, $\{y_k\}$ is also dense in $(0,A)$.

Lemma 3.5.1. For almost every sample sequence, there exist a $Q_n^* \in \mathcal{G}_n$ for each n such that

$$Q_n^* \xrightarrow{\mathcal{D}} Q_0 \quad \text{and} \quad F_{Q_n^*} \xrightarrow{\mathcal{D}} F_{Q_0} \; .$$

Proof. By the preceding lemma, we know that almost every sample sequence $\{y_k\}$ is dense in $(0,A)$. For such a sample sequence, define a sequence of distribution functions Q_n^* as follows:

Q_n^* is discrete with weights at $y_{n,1} < y_{n,2} < \cdots < y_{n,n}$

and

$$Q_n^*(y_{n,i}) = Q_0(y_{n,i}) \; , \quad i = 1,\ldots,n-1$$
$$Q_n^*(y_{n,n}) = 1 \; .$$

Since $\{y_k\}$ is dense in $(0,A)$, $Q_n^* \xrightarrow{\mathcal{D}} Q_0$. As a function of θ, $F_\theta(y)$ is continuous and bounded. So

$$F_{Q_n^*} = \int F_\theta \, dQ_n^* \xrightarrow{\mathcal{D}} F_{Q_0} = \int F_\theta \, dQ_0 \; .$$

Lemma 3.5.2. For almost every sample sequence, there exists a $Q_n^* \in \mathcal{G}_n$ for each n such that $Q_n^* \xrightarrow{\mathcal{D}} Q_0$, $F_{Q_n^*} \xrightarrow{\mathcal{D}} F_{Q_0}$ and

$$\int_0^\infty \frac{1}{\xi} \, dQ_n^*(\xi) \to \int_0^\infty \frac{1}{\xi} \, dQ_0(\xi) \; .$$

Proof. For almost every sample sequence, we claim that the same sequence Q_n^* defined in the preceding proof would also work here. The only thing that is left to be proved is

$$\int_0^\infty \frac{1}{\xi}\, dQ_n^*(\xi) \to \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi) \ .$$

The proof is by a several ε argument. Since $\int_0^\infty \frac{1}{\xi}\, dQ_0(\xi) < \infty$, by problem 2, P. 43 of Chung, there exists $a < b < c < d$, points of continuity of Q_0 such that

$$(2) \qquad \left|\int_a^d \frac{1}{\xi}\, dQ_0(\xi) - \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi)\right| < \varepsilon$$

$$(3) \qquad \left|\int_b^c \frac{1}{\xi}\, dQ_0(\xi) - \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi)\right| < 2\varepsilon \ .$$

Let

$$a_n = \min\{y_k : 1 \le k \le n,\ y_k > a\}$$

$$d_n = \max\{y_k : 1 \le k \le n,\ y_k < d\} \ .$$

Since $\{y_k\}$ is dense

$$a_n \to a \ , \qquad d_n \to d$$

$$(4) \qquad \int_0^\infty \frac{1}{\xi}\, dQ_n^*(\xi) - \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi)$$

$$= \int_0^{a_n+} \frac{1}{\xi}\, dQ_n^*(\xi) - \int_0^{a_n+} \frac{1}{\xi}\, dQ_0(\xi)$$

$$+ \int_{a_n+}^{d_n+} \frac{1}{\xi}\, dQ_n^*(\xi) - \int_{a_n+}^{d_n+} \frac{1}{\xi}\, dQ_0(\xi)$$

$$+ \int_{d_n+}^\infty \frac{1}{\xi}\, dQ_n^*(\xi) - \int_{d_n+}^\infty \frac{1}{\xi}\, dQ_0(\xi) \ .$$

Since $a_n \downarrow a$ for n sufficiently large
$d_n \uparrow d$

(5) $$\int_0^{a_n+} \frac{1}{\xi}\, dQ_n^*(\xi) < \int_0^{a_n+} \frac{1}{\xi}\, dQ_0(\xi) \quad \text{by definition of } Q_n^*$$

$$< 2\varepsilon \quad \text{by (3)} .$$

(6) $$\int_{d_n+}^{\infty} \frac{1}{\xi}\, dQ_0(\xi) < 2\varepsilon \quad \text{by (3)} .$$

(7) $$\int_{d_n+}^{\infty} \frac{1}{\xi}\, dQ_n^*(\xi) < \frac{1}{d_n} < 2\varepsilon \quad \text{without loss of generality} .$$

On the closed interval $[a,d]$, $\frac{1}{\xi}$ is uniformly continuous, so that

(8) there exists δ such that if $x,y \in [a,d]$,
$|x-y| < 2\delta$, then $\left|\frac{1}{x} - \frac{1}{y}\right| < \epsilon$.

If we divide the interval $[a,d]$ into subintervals of equal length $\delta' < \delta$, then since $\{y_k\}$ is dense in $(0,A)$

(9) for n sufficiently large, there will be some y_k among $y_1,y_2,\ldots,y_n$ lying inside each of the subintervals.

Let $y_{n,1} < y_{n,2} < \cdots < y_{n,n}$ be $y_1, y_2, \ldots, y_n$ arranged in increasing order. Define i_n, j_n by

$$a_n = y_{n,i_n} \qquad d_n = y_{n,j_n} .$$

Then

$$(10) \qquad \int_{a_n+}^{d_n+} \frac{1}{\xi}\, dQ_n^*(\xi) - \int_{a_n+}^{d_n+} \frac{1}{\xi}\, dQ_0(\xi)$$

$$= \sum_{m=i_n}^{j_n-1} \int_{y_{n,m}^+}^{y_{n,m+1}^+} \left(\frac{1}{\xi}\, dQ_n^*(\xi) - \frac{1}{\xi}\, dQ_0(\xi)\right) .$$

By definition

$$\int_{y_{n,m}^+}^{y_{n,m+1}^+} \frac{1}{\xi}\, dQ_n^*(\xi) = \frac{1}{y_{n,m+1}} \left(Q_0(y_{n,m+1}) - Q_0(y_{n,m})\right)$$

and

$$\int_{y_{n,m}^+}^{y_{n,m+1}^+} \frac{1}{\xi}\, dQ_0(\xi) = \frac{1}{y_{n,m}^*} \left(Q_0(y_{n,m+1}) - Q_0(y_{n,m})\right)$$

where $y_{n,m} < y_{n,m}^* < y_{n,m+1}$

So (10) becomes

$$(11) \qquad \int_{a_n+}^{d_n+} \frac{1}{\xi}\, dQ_n^*(\xi) - \int_{a_n+}^{d_n+} \frac{1}{\xi}\, dQ_0(\xi)$$

$$= \sum_{m=i_n}^{j_n-1} \left(\frac{1}{y_{n,m+1}} - \frac{1}{y_{n,m}^*}\right)\left(Q_0(y_{n,m+1}) - Q_0(y_{n,m})\right) .$$

For n sufficiently large

$$|y^*_{n,m} - y_{n,m+1}| < |y_{n,m} - y_{n,m+1}| < 2\delta \text{ by (9)}$$

so that

$$\left|\frac{1}{y^*_{n,m}} - \frac{1}{y_{n,m+1}}\right| < \varepsilon \quad \text{by (8)} .$$

Making use of this in (11), we get

$$\text{(12)} \qquad \int_{a_n^+}^{d_n^+} \frac{1}{\xi}\, dQ^*_n(\xi) - \int_{a_n^+}^{d_n^+} \frac{1}{\xi}\, dQ_0(\xi) < \varepsilon .$$

Apply (5), (6), (8), (12) to (4)

$$\left|\int_0^\infty \frac{1}{\xi}\, dQ^*_n(\xi) - \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi)\right| < 9\varepsilon$$

for n sufficiently large, ε is arbitrary, so

$$\int_0^\infty \frac{1}{\xi}\, dQ^*_n(\xi) \to \int_0^\infty \frac{1}{\xi}\, dQ_0(\xi) .$$

A.3. Proofs of the lemmas of section 4.6.

Lemma 4.6.1. With probability one, $\|H_n - H\| \to 0$.

Proof. This is a well known property of the sample c.d.f., see P. 133, Chung.

Lemma 4.6.2. With probability one, $\|H_{nd} - H_d\| \to 0$.

Proof. By the strong law of large numbers, for each i

$$H_{nd}(\theta_i) - H_{nd}(\theta_i -) \to H_d(\theta_i) - H_d(\theta_i -) \quad \text{a.s.} \tag{1}$$

Next, we will prove for each t, including $t = \infty$,

$$H_{nd}(t) \to H_d(t) \quad \text{a.s.} \tag{2}$$

Since

$$H_d(t) = \sum_{\substack{i \\ \theta_i \leq t}} h_i$$

given $\varepsilon > 0$, there exists $\theta_{i_1}, \theta_{i_2}, \ldots, \theta_{i_k} \leq t$, k finite such that

$$\left| \sum_{j=1}^{k} h_{i_j} - H_d(t) \right| < \varepsilon . \tag{3}$$

Since k is finite, for almost every sample sequence, $\{\theta_{i_j}\}_{j=1}^{k} \subset \Phi_n$ for n sufficiently large.

So, for almost every sample sequence,

$$\frac{1}{n} \sum_{i=1}^{n} I_{y_i \varepsilon \{\theta_{i_1}, \ldots, \theta_{i_k}\}} \leq H_{nd}(t) \leq \frac{1}{n} \sum_{i=1}^{n} I_{y_i \varepsilon \{\theta_k : \theta_k \leq t\}} \tag{4}$$

for n sufficiently large.

By the strong law of large numbers

$$\text{L.H.S. of } (4) \to \sum_{j=1}^{k} h_{i_j} \quad \text{a.s.}$$

$$\text{R.H.S. of } (4) \to \sum_{\substack{i \\ \theta_i \le t}} h_i = H_d(t) \text{ a.s.}$$

Since ε is arbitrary in (3)

$$H_{nd}(t) \to H_d(t) \quad \text{a.s.}$$

In particular,

$$\int H_{nd} \to \int H_d \quad \text{a.s.} \tag{5}$$

So if

$$\bar{H}_{nd} = \frac{H_{nd}}{\int H_{nd}}$$

$$\bar{H}_d = \frac{H_d}{\int H_d}$$

(1), (2), (5) together imply

$$\bar{H}_{nd}(\theta_i) - \bar{H}_{nd}(\theta_i -) \to \bar{H}_d(\theta_i) - \bar{H}_d(\theta_i -) \quad \text{a.s.}$$

$$\bar{H}_{nd}(t) \to \bar{H}_d(t) \quad \text{a.s.}$$

So, by the lemma on P. 133, Chung

$$\|\overline{H}_{nd} - \overline{H}_d\| \to 0 \quad \text{a.s.}$$

i.e.

$$\left\|\frac{H_{nd}}{\int H_{nd}} - \frac{H_d}{\int H_d}\right\| \to 0 \quad \text{a.s.}$$

$$\Rightarrow \|H_{nd} - H_d\| \to 0 \quad \text{a.s.} \; .$$

<u>Lemma 4.6.3</u>. With probability one, $\|H_{nc} - H_c\| \to 0$.

Proof. Direct consequence of lemma 4.6.1 and lemma 4.6.2.

<u>Lemma 4.6.4</u>. For almost every sample sequence, there exists a $Q_n^* \in \mathcal{Q}_n$ for each n such that

$$Q_n^* \xrightarrow{w} Q$$

and

$$Q_n^*(\theta_i) - Q_n^*(\theta_i-) = Q(\theta_i) - Q(\theta_i-)$$

for n sufficiently large.

Proof. The idea is essentially the same as in the proof of lemma 3.5.2. There, we show that almost every sample sequence $\{y_k\}_{k=1}^{\infty}$ is dense and we require Q_n^* to agree with Q at $y_1, y_2, \ldots, y_n$. Here, all we need to do is to add an extra requirement

$$Q_n^*(\phi_i) - Q_n^*(\phi_i-) = Q(\phi_i) - Q(\phi_i-)$$

for $\phi_i \in \Phi_n$.

Since $P(\theta_i \in \Phi_\infty) = 1$ by (4.5.3). For almost every sample sequence, $\theta_i \in \Phi_n$ for n sufficiently large so that

$$(6) \qquad Q_n^*(\theta_i) - Q_n^*(\theta_i -) = Q(\theta_i) - Q(\theta_i -)$$

for n sufficiently large.

<u>Lemma 4.6.5</u>. For almost every sample sequence, let Q_n^* be defined as in lemma 4.6.4, $\|F_{Q_n^*} - F_Q\| \to 0$.

Proof. (4.3.2) and (6) together imply

$$F_{Q_n^*}(\theta_i) - F_{Q_n^*}(\theta_i -) \to F_Q(\theta_i) - F_Q(\theta_i -) \ .$$

As a function of y, $F_\theta(y)$ has a jump at $\theta = y$, it is otherwise continuous and $\lim_{\theta \to \infty} F_\theta(y) = 0$. So if y is a point of continuity of F_Q, by a slight extension of Theorem 4.4.1, P. 87, Chung

$$F_{Q_n^*}(y) = \int F_\theta(y)\, dQ_n^*(\theta) \to \int F_\theta(y)\, dQ(\theta) = F_Q(y) \ .$$

Again, by the lemma on P. 133, Chung

$$\|F_{Q_n^*} - F_Q\| \to 0 \ .$$

<u>Lemma 4.6.6</u>. For almost every sample sequence, let Q_n^* be as defined in lemma 4.6.4.

$$\|F^{(n)}_{Q_n^*,d} - H_d\| \to 0 \ .$$

Note: See section 4.3d for definition of $F^{(n)}_{Q^*_n,d}$.

Proof. From (4.3.2)

$$H_d = \sum_{i=1}^{m} \frac{\tau}{\tau+\theta_i} q_i \delta_{\theta_i} \quad . \tag{7}$$

By construction, Q^*_n puts mass q_i at θ_i if $n_i > 1$ so that (4.3.3) becomes

$$F^{(n)}_{Q^*_n,d} = \sum_{i=1}^{m} \frac{\tau}{\tau+\theta_i} q_i \, 1_{\{n_i > 1\}} \, \delta_{\theta_i} \tag{8}$$

$$\frac{n_i}{n} \to h_i = \frac{\tau}{\tau+\theta_i} q_i \Rightarrow n_i > 1 \quad \text{for } n \text{ sufficiently large} .$$

A comparison of (7) and (8) then tells us

$$F^{(n)}_{Q^*_n,d}(\theta_i) - F^{(n)}_{Q^*_n,d}(\theta_i -) = H_d(\theta_i) - H_d(\theta_i -) \tag{9}$$

$$= \frac{\tau}{\tau+\theta_i} q_i \quad \text{for } n \text{ sufficiently large.}$$

Given t,

$$H_d(t) = \sum_{\substack{i \\ \theta_i \le t}} \frac{\tau}{\tau+\theta_i} q_i \quad .$$

Given $\varepsilon > 0$, there exists $\theta_{i_1}, \ldots, \theta_{i_k} \le t$, k finite such that

$$\left| \sum_{j=1}^{k} \frac{\tau}{\tau+\theta_{i_j}} q_{i_j} - H_d(t) \right| < \varepsilon \quad . \tag{10}$$

Since k is finite, for n sufficiently large $n_{i_j} > 1$, $j = 1,\ldots,k$, so that

$$\sum_{j=1}^{k} \frac{\tau}{\tau+\theta_{i_j}} q_{i_j} \leq F^{(n)}_{Q^*_n,d}(t) \leq H_d(t) .$$

Since ε is arbitrary in (10)

$$F^{(n)}_{Q^*_n,d}(t) \to H_d(t) .$$

This and (9) imply

$$\| F^{(n)}_{Q^*_n,d} - H_d \| \to 0 .$$

<u>Lemma 4.6.7</u>. For almost every sample sequence, let Q^*_n be as defined in lemma 4.6.4,

$$\| F^{(n)}_{Q^*_n,c} - H_c \| \to 0 .$$

Proof. Direct consequence of lemma 4.6.5 and lemma 4.6.6.

<u>Lemma 4.6.8</u>. If $F_{Q_n} \xrightarrow{\mathcal{D}} F_Q$, then $Q_n \xrightarrow{\mathcal{D}} Q$.

Proof. Since $F_{Q_n} \xrightarrow{\mathcal{D}} F_Q$, if t is a point of continuity of F_Q,

(11) $$F_{Q_n}(t) = \int_0^\infty F_\theta(t)dQ_n(\theta) \to F_Q(t) = \int_0^\infty F_\theta(t)dQ(\theta) .$$

Let Q_{n_i} be any subsequence of Q_n such that $Q_{n_i} \xrightarrow{\mathcal{D}} Q^*$ where Q^* is a subdistribution function, we will be done if we can show $Q^* = Q$. Suppose t is also a point of continuity of Q^*, then as in the proof of lemma 4.6.5, by extending Theorem 4.4.1, P. 87, Chung slightly

$$\int_0^\infty F_\theta(t)\ dQ_{n_i}(\theta) \to \int_0^\infty F_\theta(t)\ dQ^*(\theta)\ .$$

Comparison with (11) yields

$$(12) \qquad \int_0^\infty F_\theta(t)\ dQ(\theta) = \int_0^\infty F_\theta(t)\ dQ^*(\theta)\ .$$

Since $\lim_{t\to\infty} F_\theta(t) = 1$, by monotone convergence

$$\begin{aligned}\int_0^\infty dQ(\theta) &= \lim_{t\to\infty} \int_0^\infty F_\theta(t) dQ(\theta) \\ &= \lim_{t\to\infty} \int_0^\infty F_\theta(t) dQ^*(\theta) \\ &= \int_0^\infty dQ^*(\theta)\ .\end{aligned}$$

So Q^* is a distribution function. Since (12) holds whenever t is a point of continuity of both F_Q and Q^* and the set of such points is dense, $F_Q = F_{Q^*}$, so $Q = Q^*$ by identifiability. Therefore $Q_n \xrightarrow{\mathcal{D}} Q$.

Lemma 4.6.9. If G is discrete, then for almost every sample sequence, there exist Q_n^*, Q_n^* discrete, puts weights at $\phi_1, \phi_2, \ldots, \phi_{m(n)}$ such that the statements of lemmas 4.6.4, 4.6.5, 4.6.6, 4.6.7 still hold.

Proof. The proof is similar to the proofs of lemmas 4.6.4, 4.6.5, 4.6.6, and 4.6.7 and will be omitted.

REFERENCES

Anderssen, R.S. and Bloomfield, P. (1973). Numerical Differentiation Procedures for Non-Exact Data. Num. Math., 22, 157-182.

Anderssen, R.S. and Jakeman, A.J. (1975). Product Integration for Functionals of Particle Size Distributions. Util. Math., 8, 111-126.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D. (1972). Statistical Inference Under Order Restrictions. Wiley, New York.

Bloomfield, P. On Estimating the Mean of a Distribution with Infinite Variance. (unpublished manuscript).

Choi, Keewhan and Bulgren, W.G. (1968). An Estimation Procedure for Mixture of Distributions. JRSS B, 30, 444-460.

Chung, K.L. (1968). A Course in Probability Theory. Academic Press.

Coleman, R. (1979). An Introduction to Mathematical Stereology. Memoir Series. Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, Denmark.

Deely, J.J. and Kruse, R.L. (1968). Construction of Sequences Estimating the Mixing Distribution. Ann. Math. Statist., 39, 286-288.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. JRSS B, 39, 1-38.

Jakeman, A.J. and Anderssen, R.S. (1976). On Optimal Forms of Stereological Data. Pages 69-74 in Underwood et. al (1976).

Jakeman, A.J. and Scheaffer, R.L. (1978). On the Properties of Product Integration Estimators for Linear Functionals of Particle Size Distributions. Util. Math., 14, 117-128.

Mecke, J. and Stoyan, D. (1980). Stereological Problems for Spherical Particles. Math. Nachr., 96, 311-317.

Moran, P.A.P. (1972). Probabilistic Basis of Stereology. Pages 69-91 in Nicholson (1972).

Nicholson, W.L. (1970). Estimation of Linear Properties of Particle Size Distributions. Biometrika, 57, 273-297.

Nicholson, W.L. (1972). Proceedings of the Symposium on Statistical and Probabilistic Problems in Metallurgy. Seattle, Washington. Special Supplement to Advanced Applied Probability (1972).

Robbins, Herbert (1964). The Empirical Bayes Approach to Statistical Design Problems. Ann. Math. Statist., 35, 1-20.

Stoyan, D. (1979). On Some Qualitative Properties of Boolean Model in Stochastic Geometry.

Underwood, E.E., de Wit, R. and Moore, G.A. (Editors)(1976). Proceedings of the Fourth International Congress for Stereology. (Gaithersburg, 1975). National Bureau of Standard Special Publications 431.

Watson, G.S. (1971). Estimating Functionals of Particle Size Distributions. Biometrika, 58, 483-490.

Wegman, E.J. (1970a). Maximum Likelihood Estimation of a Unimodal Density Function. Ann. Math. Statist., 41, 457-471.

Wegman, E.J. (1970b). Maximum Likelihood Estimation of a Unimodal Density, II. Ann. Math. Statist., 41, 2169-2174.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 328	2. GOVT ACCESSION NO. AD-A120 687	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A MIXING DISTRIBUTION APPROACH TO ESTIMATING PARTICLE SIZE DISTRIBUTIONS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) ANTHONY YUNG C. KUK		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office Of Naval Research Statistics & Probability Program Code 411SP Arlington, VA 22217		12. REPORT DATE OCTOBER 19, 1982
		13. NUMBER OF PAGES 123
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stereology, Particle size distributions, Nonparametric estimation of mixing distributions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Spherical particles are dispersed randomly in a three-dimensional body. The centers of the spheres are distributed according to a dilute Poisson process. The radii of such spheres have a distribution G independent of everything else. A random probe (line, plane or thin

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

slice) is cut through the volumes. The observations y are the following: i) For the random line, its intersection with a sphere is a line segment, y is half the length of that segment. ii) For the random plane, y is the radius of the circle of intersection. iii) For the thin slice, y is the maximum of the radii of the circles of intersection. The problem is to estimate G from the observations $y_1,\ldots,y_n$.

Past procedures make use of the inversion formula that expresses the particle size distribution G as a function of the distribution H of the observed data. An estimate $\tilde{G}$ is obtained by replacing H in the formula by an estimate $\tilde{H}$ where $\tilde{H}$ is either the sample c.d.f. or a smooth version of it. These procedures do not take the structure of the problem into account. Consequently, they have some serious shortcomings. Taking the viewpoint of nonparametric estimation of mixing distributions, we propose a new procedure that deals with the shortcomings of the classical procedures. We consider linear, planar and thin slice data. In all three cases, our approach performs better than the classical procedure. In addition, we prove consistency results.

In the random plane case, we discuss the right way and the wrong way to bootstrap the distribution of a stereological estimate, corresponding to whether we have taken the structure of the problem into account or not. In the thin slice case, when G is mixed or discrete, the formulas involve a decomposition of H into its continuous and discrete component. This makes the estimation problem more complicated but also more interesting especially in the discrete case. We propose a few procedures which involve a decomposition of the data corresponding to that of H.